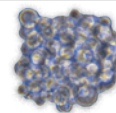


THIS WEEK

EDITORIALS

NUTRITION There is more to the food debate than organic claims **p.264**

WORLD VIEW Facebook experiment has been unfairly criticized **p.265**



CANCER Tumour cells harvested from blood and cultured **p.266**

Struggle for independence

The faculty of the Scripps Research Institute is bucking a national trend with its refusal to merge with the University of Southern California.

Remarkable events are unfolding at the Scripps Research Institute in La Jolla, California, where faculty members have rebelled against their president's attempts to merge with the University of Southern California (USC) in Los Angeles. These struggles are emblematic of today's difficult landscape for independent biomedical research institutes in the United States. Highly dependent on funding from the US National Institutes of Health (NIH), many independents have closed or merged with larger institutions (see *Nature* **491**, 510; 2012), and Scripps president Michael Marletta wanted his centre to join that trend. In June, news leaked that Marletta had brokered a potential deal that would have seen USC pay Scripps US\$600 million so that the two institutions could join up.

But in an interesting departure from the script, Scripps faculty members have said no to the deal, have argued against its entire basis and have now taken matters into their own hands. As we report on page 274, they have passed a vote of no confidence in Marletta by a startling margin — almost unanimously. They say that they can solve Scripps' financial crisis without his help, thank you very much, and can do so without selling out the institution that they love. Are they right? Other labs are watching with interest.

The impasse is a product of clashing views on Scripps, a prestigious independent institute that regularly attracts more than \$300 million a year in NIH funding — upwards of 80% of the institute's operating budget. A sizeable chunk of the rest has tended to come from the pharmaceutical industry, but that has declined sharply in recent years, leaving the institute with a projected \$21-million budget gap for this fiscal year.

But where Marletta sees this deficit as a problem necessitating a change in how Scripps does business, faculty members claim that it is a temporary setback, not an existential threat, and one that should be solved without changing the nature of their institute. They fear that a merger with USC would compromise their cherished independence — many point out that although they would get more job security at larger institutions, they have chosen to work at Scripps because its small size and non-hierarchical nature free them from administrative burdens and teaching that would distract them from science. And they are angry at Marletta's decision to negotiate the USC deal in secret, feeling that as Scripps' main breadwinners, they deserved to know much earlier that he was even considering such a move.

The closed-door negotiations have raised suspicions among faculty members that Marletta does not understand their priorities — or worse, that he does not share them. They think that the \$600 million he agreed to, which was to be meted out in \$15-million increments over 40 years, was a vast undervaluing of Scripps assets, including its formidable grant money, sizeable investments and coveted seaside location. To many, the deal felt like a land grab by USC, which would have paid a bargain rate for scientific prestige, a valuable piece of land

and a southern foothold for its health-care practice.

The whole episode has cemented the faculty members' growing mistrust of Marletta, who has been president of Scripps since January 2012; previously, he was chair of the chemistry department at the University of California, Berkeley. Many at Scripps, including Marletta himself, feel that philanthropy could plug the institution's budget gap. But the president has brought in no major donations during his term. By contrast, the Sanford–Burnham and Salk biomedical-research institutes, also in La Jolla, have each raised hundreds of millions of dollars in recent years. Scripps faculty members say that there is clearly donor money available in their wealthy area, and Scripps could do more to access it, perhaps by accentuating its strengths in chemistry and chemical biology.

How the institution will get itself out of this situation is not clear. The faculty members think that they can find a way to close the budget gap by themselves and are determined to try. It would certainly prove a coup. But they would also benefit from having a full-time leader whose entire job is focused on their future.

Whether Marletta is this person is currently up for debate. It would probably be in the best interests of everyone at Scripps if he could find a way to demonstrate to the faculty members that he has heard their concerns and will change his approach. If he can do that, Scripps will be more likely to buck the trend of small institutes succumbing to their budget woes. ■

“Faculty members think that they can find a way to close the budget gap by themselves, and are determined to try.”

Within reach

A redoubling of efforts should swiftly eradicate polio from its last strongholds.

The global effort to eradicate poliomyelitis has been spectacularly successful, eliminating 99% of cases in its 26-year history. But that progress has begun to unravel in the past 18 months, with outbreaks in east and west Africa and in the Middle East. The lesson is clear: as long as the virus is allowed to persist in the three countries in which it remains endemic — Pakistan, Afghanistan and Nigeria — exports of the disease will continue to affect other countries. A determined effort is needed to eradicate the virus from these endemic countries, and fast.

The worsening situation meant that in May, the World Health Organization (WHO) declared polio a public-health emergency of

international concern. This allowed it to impose a requirement that all travellers entering or leaving Pakistan, Cameroon, Syria and Equatorial Guinea — the countries currently exporting polio — must have up-to-date polio vaccinations. And it strongly recommended the same for other nations with ongoing polio outbreaks. The WHO also requires the governments of affected countries to declare that polio constitutes a national public-health emergency.

It is too soon to tell how well countries will enforce the travel restrictions or how effective they will be (see page 285). But the WHO's declaration has another, and arguably more important, potential impact. It has greatly heightened public and political awareness of the global polio threat. The move could yet shame those nations with weak control efforts into doing better. Ultimately, political will, through every level of government right down to the local level, is crucial if eradication efforts are to succeed.

The setbacks have reignited scepticism among some critics of the multibillion-dollar global effort, which has repeatedly missed its own deadlines for worldwide eradication — the first such deadline was set for 2000. But this must not obscure the fact that impressive gains have been made, so much so that at the end of 2012, global polio eradication truly seemed within reach. It is important to turn the current situation around quickly, consolidate those gains, and condemn polio to the history books.

There is cause for optimism. In Afghanistan, the virus has been wiped out from many areas where it was previously rampant, with cases now restricted mostly to the northeast, where polio is imported from across the border with Pakistan. Afghanistan is expected to become polio-free perhaps as soon as year's end. Nigeria has also improved its eradication efforts, resulting in a sharp drop in case numbers. Eradication there is in sight, although a current worsening of the country's political and security tensions risks undoing the progress. Pakistan, despite a lacklustre control effort, has also shrunk the geographical range of the virus.

The global-eradication effort — despite some shortcomings — has a good track record of successfully fighting sporadic flare-ups. There is every reason to believe that the current spate of outbreaks will be contained (although war-torn Syria could remain problematic).

The big challenge is to conquer the virus in the endemic countries that are fuelling exports of the disease — and above all in Pakistan. A report released in May by the Independent Monitoring Board of the Global Polio Eradication Initiative puts it bluntly: "Pakistan's situation is dire. Its program is years behind the other endemic countries." Unless matters change, the report concludes, the country is "firmly on track to be the last polio-endemic country in the world".

That damning indictment needs to be heard and responded to at every level of Pakistani society. The country faces many obstacles — but so too did the other countries that nonetheless have succeeded in eradicating polio. There is no excuse for Pakistan not to do so. Its government must pull out all the stops to act swiftly and decisively. As the report rightfully argues, ultimate responsibility for Pakistan's bungled polio efforts lies with its authorities: "If the country's leaders were to truly and wholly take on the mission of wiping polio from their borders, what now seems to some an impossible dream would fast become reality."

Another barrier to eradication is societal resistance to vaccination, rooted, for example, in local distrust of immunization campaigns and unfounded concerns that it conflicts with religious beliefs. Polio has spread to Waziristan in northern Pakistan, a stronghold of the Taliban, who have banned vaccinations. Vaccinators have also been murdered.

In the past few months, international Islamic scholars and bodies — including the newly formed Islamic Advisory Group on Polio Eradication — have to their credit spoken out to condemn attacks on polio workers, and to emphasize that polio vaccination is compatible with Islam, denouncing those who claim otherwise. Resistance and suspicion of vaccines will always be present, but religious leaders can help by reiterating these messages to local populations.

Pakistan's situation is exacerbated by the Taliban's stubborn blocking of polio vaccinations, ostensibly in opposition to US drone strikes. But polio has no religion. It respects no political affiliation. For the benefit of all, every effort must be made to overcome residual resistance to vaccination and to root out the virus from its last strongholds. ■

"Ultimate responsibility for Pakistan's bungled polio efforts lies with its authorities."

Food for thought

Researchers investigating different farming practices should not have to pick sides.

Some debates run and run. Last month, an analysis found that a selection of organically farmed food contained, on average, higher concentrations of supposedly beneficial antioxidant compounds than food produced by conventional farming (M. Barański *et al.* *Br. J. Nutr.* <http://doi.org/tqs; 2014>).

This field is still relatively small and the quality of research can be variable. The analysis advances previous work, thoroughly evaluates the current situation and yields some results that warrant further investigation. Still, several prominent nutrition scientists have voiced valid criticisms of the paper's method and statistical analysis (see go.nature.com/ikx15z), and have raised concerns over the scientific rigour of some of the primary research that it covers.

It is good to be thorough: the study examines all of the available evidence so far. But in a field in which research quality can be hit and miss, it can be better to be cautious. The authors would perhaps have generated more confidence in their results if they had been more selective. But such selection, inevitably, raises questions about how it is done.

Beyond the arguments about this specific study, which the authors have defended, lies a bigger issue. There are some fundamental

questions that this type of research cannot answer, despite the way the results have been interpreted by the mainstream media as pointing to clear benefits of organic farming.

The study attempts to examine how different farming methods affect the nutritional quality of the product — an important question. There is plenty of room for improvement in the conventional farming system and in the nutritional quality of many people's diets. So far, so good.

The paper also refers to the link between increased dietary concentrations of antioxidant compounds, such as phenolic acids and flavonols, and a reduced risk of chronic diseases — including some cancers. However, the evidence for such a link is mixed, and tentative at best. A more important question is not the level of antioxidants in organic or non-organic food, but how that contributes to health.

It is also not clear that organic farming practices are the cause of the observed higher concentrations of antioxidants. Research could help to determine, for example, whether organic crops — which are not treated with pesticides — release more phenolic compounds as a defence against pests and pathogens. Or perhaps the nitrogen fertilizers applied to conventional crops encourage growth rather than the production of such chemical defences.

This is a useful discussion, but difficult to have on neutral territory. Research on the different farming systems can often seem like a contest in which one practice is pitted against another and in which researchers must pick sides. Science should stay focused on the heart of the matter: the provision of more nutritious food for more people in a more sustainable way. ■

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhunq

DAVID SHAFFER



Misjudgements will drive social trials underground

A Facebook study that manipulated news feeds was not definitively unethical and offered valuable insight into social behaviour, says **Michelle Meyer**.

Some bioethicists have said that Facebook's recent study of user behaviour is "scandalous", "violates accepted research ethics" and "should never have been performed".

I write with 5 co-authors, on behalf of 27 other ethicists, to disagree with these sweeping condemnations (see go.nature.com/my5lvz).

We are making this stand because the vitriolic criticism of this study could have a chilling effect on valuable research. Worse, it perpetuates the presumption that research is dangerous.

When the average user logs on, Facebook automatically chooses 300 status updates from a possible 1,500 to display in his or her feed. Such manipulation, which often determines how likely people are to view emotionally charged content, aims to optimize user engagement and activity and is how Facebook is able to offer a free service but still make a profit. But how does this affect users' moods?

No one knows whether exposure to a stream of baby announcements, job promotions and humble brags makes Facebook's one billion users sadder or happier. The exposure is a social experiment in which users become guinea pigs, but the effects will not be known unless they are studied.

For a week in January 2012, a data scientist from Facebook, along with two researchers from Cornell University in Ithaca, New York, tried to do just that. Of the many millions of users who log on every day, they randomly selected 310,000. Automated software — not researchers who read users' feeds, as some have suggested — coded a post as 'positive' or 'negative' if it contained a single such word.

Facebook then adjusted its algorithm to filter from half of these feeds 10–90% of the positive content, and from the other half a similar amount of negative content (A. D. I. Kramer, J. E. Guillory and J. T. Hancock *Proc. Natl Acad. Sci USA* **111**, 8788–8790; 2014). This had the effect of concentrating the feeds with negative and positive content, respectively.

Some have said that Facebook "purposefully messed with people's minds". Maybe; but no more so than usual. The study did not violate anyone's privacy, and attempting to improve users' experience is consistent with Facebook's relationship with its consumers.

It is true that Facebook altered its algorithm for the study, but it does that all the time, and this alteration was not known at the time to increase risk to anyone involved. Academic studies have suggested that users are made unhappy by exposure to positive posts (E. Kross *et al. PLoS ONE* **8**, e69841; 2013). The results of Facebook's study pointed in the opposite direction: users who were exposed to less positive content very slightly decreased their own use of positive words and increased their use of negative words.

We do not know whether that is because

negativity is 'contagious' or because the complaints of others give us permission to chime in with the negative emotions we already feel. The first explanation hints at a public-health concern. The second reinforces our knowledge that human behaviour is shaped by social norms. To determine which hypothesis is more likely, Facebook and academic collaborators should do more studies. But the extreme response to this study, some of which seems to have been made without full understanding of what it entailed or what legal and ethical standards require, could result in such research being done in secret or not at all.

Let us be clear. If critics think that the manipulation of emotional content in this research is sufficiently concerning to merit regulation or charges of unethical behaviour, then the same concern must apply to Facebook's standard practice — and many similar practices by companies, non-profit organizations and governments.

But if it is ethically permissible for Facebook to offer a service that carries unknown emotional risks, and to alter that service to improve user experience, then it should be allowed — and encouraged — to try to quantify those risks and publish the results.

Much has been made of the issue of informed consent, which the researchers did not obtain. Here, there is some disagreement even among the six of us. Some think that the procedures were consistent with users' reasonable expectations of Facebook and that no explicit consent was required. Others argue that the research imposed little or no incremental risk and that informed consent might have biased the results; in those circumstances, ethical guidelines,

such as the US regulations for research involving humans, permits researchers to forgo or at least substantially alter the elements of informed consent.

Although approval by an institutional review board was not legally required for this study, it would have been better for everyone involved had the researchers sought ethics review and debriefed participants afterwards.

The Facebook experiment was controversial, but it was not an egregious breach of either ethics or law. Rigorous science helps to generate information that we need to understand our world, how it affects us and how our activities affect others. Permitting Facebook and other companies to mine our data and study our behaviour for personal profit, but penalizing it for making its data available for others to see and to learn from makes no one better off. ■

THE EXTREME
RESPONSE
TO THIS STUDY
COULD RESULT IN
SUCH RESEARCH
BEING DONE IN
SECRET.

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/my5lvz

Michelle N. Meyer is director of bioethics policy at the Union Graduate College-Icahn School of Medicine at Mount Sinai Bioethics Program in New York.
e-mail: michellenmeyer@gmail.com

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

PLANETARY SCIENCE

Titan's sea is super salty

Saturn's largest moon, Titan, has a buried ocean that is saltier than many seas on Earth.

Titan, with its thick atmosphere and bodies of surface liquid, is of great interest to scientists looking for life beyond Earth. A team led by Giuseppe Mitri, of the National Institute of Astrophysics in Rome, looked at gravity and elevation measurements taken by NASA's Cassini spacecraft over more than a decade.

The scientists calculated that Titan's icy outer shell is less than 100 kilometres thick and is in the process of freezing and growing thicker. They also calculated that the underlying water is about as dense as the Dead Sea, probably because of high concentrations of sulphur, potassium, sodium and other salts, the authors say.

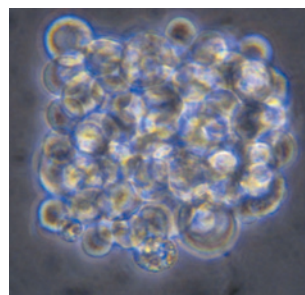
Icarus 236, 169–177 (2014)

CANCER

Roving tumour cells tracked down

Cancer cells in the blood can now be isolated and studied in culture, opening up the possibility of personalizing treatment strategies.

Tumours shed small amounts of cancer cells into



the bloodstream, but it has been difficult to isolate and grow these cells. Shyamala Maheswaran and Daniel Haber of Massachusetts General Hospital in Boston and their colleagues developed an improved microfluidic system that filters out normal blood cells, leaving tumour cells unharmed.

The team used the device to harvest circulating tumour cells from the blood of patients with advanced breast cancer. These were then grown in culture (**pictured**) and sequenced to reveal key mutations in certain cancer genes. The researchers also

tested the cells' sensitivity to various drugs.

With further improvements, the technique could one day be used to guide therapy, the authors say.

Science 345, 216–220 (2014)

ECOLOGY

Ocean reserves miss key target

Marine reserves may not be protecting the world's most vulnerable reef-fish communities.

Marine protected areas exist mainly in regions with a large number of different fish species. Valeriano

Parravicini at the Centre for the Synthesis and Analysis of Biodiversity in Aix-en-Provence, France, and his colleagues mapped the ranges of more than 6,000 species of tropical reef fishes and quantified the sensitivity of these species to human threats.

They found that areas where species are vulnerable to extinction do not often overlap with protected regions of high species richness. For example, seas off the coast of Chile and the eastern Atlantic were areas of high vulnerability, but species-rich hotspots are centred around Indonesia and Australia.



AGRICULTURE

Global warming could hurt crops

The warming climate could put food supplies at risk over the next decade or two.

Using various combinations of climate models, David Lobell at Stanford University, California, and Claudia Tebaldi at the National Center for Atmospheric Research in Boulder, Colorado, compared expected yields of maize (corn) and wheat growing under natural

climate variations to projected yields influenced by human-induced climate change. The results suggest that with climate warming, the risk of losing 10% or more of the global wheat yield over the next two decades increases tenfold, to a 1 in 20 chance. For maize, the risk increases by 20 times, to a 1 in 10 chance.

Environ. Res. Lett. 9, 074003 (2014)

STEVE CAVALIER/ALAMY

MIN YU

More marine areas need to be protected to maintain tropical fish biodiversity, the authors say.

Ecol. Lett. <http://doi.org/tn4> (2014)

NEURODEGENERATION

Antibodies fight Parkinson's

Antibodies that target a protein associated with Parkinson's disease reverse some symptoms in a mouse model of the neurodegenerative disorder.

In the brains of patients with Parkinson's, the α -synuclein protein clumps together and spreads between cells. Eliezer Masliah at the University of California, San Diego, and his colleagues made various antibodies that bind to one end of the protein, and injected them into transgenic mice that overexpress α -synuclein. Some of the antibodies reduced the accumulation of α -synuclein in the animals, improved their memory and movement, and, in cell culture, reduced the spreading of α -synuclein between cells.

By binding to one end of α -synuclein, the antibodies prevent the protein from aggregating and propagating, the authors suggest.

J. Neurosci. 34, 9441–9454 (2014)

APPLIED PHYSICS

Phone powers electronic label

A small electronic device that is powered by wireless signals from mobile phones could one day be used to label and connect a wide range of products to the Internet.

A team led by Magnus

Berggren at Linköping University and Göran Gustafsson at Acreo Swedish ICT, both in Norrköping, Sweden, developed a printed, flexible silicon diode with a small antenna that picks up the signal emitted by a nearby phone during a call. The diode then converts the signal to a current that powers a display (pictured).

Electronic labels that can communicate with web-connected devices could be important for a future 'Internet of things', in which ubiquitous objects such as sensors and appliances can be controlled through the Internet.

Proc. Natl Acad. Sci. USA <http://doi.org/tnz> (2014)

VISION

Prism of the eye guides light

A group of cells in the retina splits white light and channels specific wavelengths to light sensors to improve daytime vision.

Amichai Labin, Ido Perlman and their colleagues at the Technion Israel Institute of Technology in Haifa used a computer model to study the role of Müller cells, which funnel light towards light-sensitive cells in the human retina.

The team found that Müller cells concentrate green and red light onto the daytime-light-sensing cones, increasing by up to ten times the amount of light they absorb than if Müller cells were absent. Blue light, however, leaks out of Müller cells towards rod cells, which enable night vision. Imaging experiments on isolated guinea-pig retinas largely confirmed the model's results.

The findings could explain how light is able to travel efficiently through various cellular layers in the retina to reach the cone cells. *Nature Commun.* 5, 4319 (2014)

SOCIAL SELECTION

Popular articles on social media

Bigfoot sighted on Twitter

Researchers had some fun on social media with a rare appearance of Bigfoot in the scientific literature. A team led by geneticist Bryan Sykes at the University of Oxford, UK, ran DNA tests on 30 hair samples reputed to come from "anomalous" primates, including Bigfoot and the Himalayan yeti. As it turned out, the origins of the hairs could be explained without invoking any elusive hominins. Malcolm Campbell, a cell biologist at the University of Toronto, summed up the paper in his tweet: "Cows, and horses, and bears, oh my. 'Bigfoot' & 'Sasquatch' samples come from extant mammals." And plant scientist David Baltrus of the University of Arizona in Tucson tweeted: "That clump of Bigfoot hair you found outside your cabin ... yeah, prolly not Bigfoot."

Proc. R. Soc. B. 281, 20140161 (2014)



Based on data from altmetric.com. Altmetric is supported by Macmillan Science and Education, which owns Nature Publishing Group.

➔ **NATURE.COM**
For more on popular papers:
go.nature.com/nqtojc

VIROLOGY

What makes HIV fit to spread

HIV isolated from newly infected people tends to have certain genetic variations that help it to thrive in its new host.

When HIV-1 spreads from one heterosexual partner to another, a single viral variant typically takes hold. To determine if these successful viruses share any traits, a team led by Jonathan Carlson at Microsoft Research in Redmond, Washington, and Eric Hunter at Emory University in Atlanta, Georgia, analysed viral genetic diversity in 137 heterosexual pairs shortly after HIV transmission from one partner to the other.

The viruses that established infection tended to have the same genetic mutations that boost fitness — for instance by improving the stability of the virus's proteins.

Drugs or vaccines that drive the selection of even slightly less fit HIV variants could prevent new infections, even

when the virus is transmitted, the researchers say.

Science <http://doi.org/tpc> (2014)

COGNITION

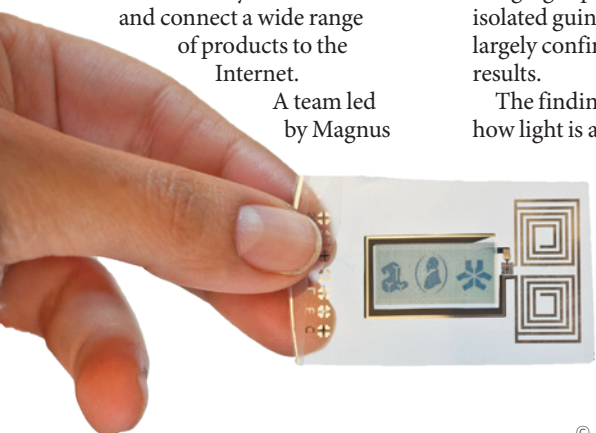
Chimp intelligence partly inherited

Genetics could explain about half of the intelligence of chimpanzees.

William Hopkins and his colleagues at Georgia State University in Atlanta used a battery of tests to measure the intelligence of 99 chimpanzees aged 9 to 54 years old. A statistical analysis revealed a correlation between the animals' performance on these tests and their relatedness to other chimpanzees in the study. About half of the difference in performance between individual apes was genetic.

The findings could lead to the discovery of genes linked to intelligence, the authors say. *Curr. Biol.* <http://doi.org/tn3> (2014)

➔ **NATURE.COM**
For the latest research published by Nature visit:
www.nature.com/latestresearch



SEVEN DAYS

The news in brief

RESEARCH

Brain-project reply

The leaders of the European Union's Human Brain Project (HBP), which was launched last October, responded on 9 July to an open letter from hundreds of neuroscientists who criticized the project's scientific plans and management. The 7 July letter was addressed to the European Commission, which co-funds the HBP, worth €1 billion (US\$1.4 billion) over 10 years. In a statement on its website, the project's executive committee acknowledged the signatories' concerns and said that the HBP's governance would evolve as it developed. The commission promised to engage with all parties. See go.nature.com/ttydmb for more.

Faked peer review

On 8 July, the publisher SAGE announced that it had retracted 60 articles after a 14-month investigation into what it termed a "peer review and citation ring" at the *Journal of Vibration and Control*. The publisher said that fabricated identities had been used on its online manuscript-submission system. It specifically identified the "strongly suspected misconduct" of Peter Chen, formerly of the National Pingtung University of Education in Taiwan. Taiwanese education minister Chiang Wei-ling resigned on 14 July because his name appeared as a co-author on some of the retracted articles.

HIV-cure rethink

Paediatricians announced on 10 July that HIV has resurfaced in the little girl known as the 'Mississippi baby', who was thought to have been cured. The girl, now almost four, was treated with antiretroviral drugs



ANDREI MORGUNOV/RUSSIAN MINISTRY OF DEFENCE/ITAR-TASS/ALAMY

Russian rocket makes maiden flight

After repeated delays, Russia has successfully test-launched Angara, its first rocket developed after the dissolution of the Soviet Union. The rocket blasted off from the Plesetsk Cosmodrome in northwestern Russia on 9 July just after 12.00 GMT. In development for

more than two decades, the next-generation Angara rockets are intended to ease reliance on technologies of former Soviet countries. Russian military news agency Interfax-AVN reported that parts of the rocket had, as planned, fallen in the Kamchatka Peninsula in eastern Russia.

from soon after birth until she reached 18 months of age, after being born to an HIV-positive mother who had not received treatment during her pregnancy. A federally funded trial aiming to replicate the results is now being reevaluated. See go.nature.com/2ak8lg for more.

and storage scheme to win support, although the fund originally intended to reward more (see *Nature* **493**, 141–142; 2013). The rest of the money went to wind, solar and other renewable-energy projects. The commission confirmed the winners on 8 July.

is preserved in only two repositories worldwide. NIH safety officials determined that the virus had not leaked and there was no danger to the employees who had found it. See go.nature.com/tr4ehk for more.

Exoplanet names

The International Astronomical Union (IAU) wants to bolster its role in naming celestial bodies. The IAU announced on 9 July that it will partner with the citizen-science group Zooniverse to solicit popular exoplanet names from astronomy clubs and non-profit groups. The public will then be able to vote on names through a web platform. Last year, the space-education company Uwingu in Boulder, Colorado,

FUNDING

Carbon burial

A €2-billion (US\$2.7-billion) European Commission fund for low-carbon energy will give €300 million from its second funding round to a UK coal-power plant that plans to bury its carbon dioxide emissions under the North Sea. The White Rose project near Selby, planned to be operational by 2018, is the only carbon capture

EVENTS

Smallpox found

The US Centers for Disease Control and Prevention announced on 8 July that forgotten stores of the potentially deadly smallpox virus had been discovered in a refrigerator belonging to the US Food and Drug Administration on the National Institutes of Health (NIH) campus in Bethesda, Maryland. Officially, smallpox

ran a competition in which the public could pay to suggest and vote on popular names for exoplanets, most of which currently have technical designations (see *Nature* 496, 407; 2013).

Out of gas

A citizen-science effort attempting to retrieve NASA's decades-old International Sun-Earth Explorer-3 has hit a snag. The group was trying to shift the craft's trajectory on 8 July when the thrusters ran out of the nitrogen gas used to pump hydrazine fuel. The probe, launched in 1978, continues to transmit data that the team intends to monitor. See go.nature.com/rgvhtq for more.

FACILITIES

Conflict avoidance

The California Institute for Regenerative Medicine (CIRM) has told its employees and board members not to discuss business with its former president, Alan Trounson, to avoid any conflicts of interest. After departing CIRM, on 7 July, Trounson (pictured) joined the board of StemCells in Newark, California, which has been awarded \$19.4 million by CIRM. In a 9 July statement, CIRM outlined the state-imposed restrictions on Trounson's interactions with



the agency, and said that it will review past activities with the company in the interests of transparency.

Dark-matter money

Three next-generation dark-matter projects have won the backing of the US Department of Energy and the National Science Foundation. On 11 July, the agencies announced that they would fund the Super Cryogenic Dark Matter Search–SNOLAB Experiment, to be built near Sudbury in Canada, and the LUX ZEPLIN collaboration at the Sanford Underground Research Facility in South Dakota. They will also support an upgrade to the Axion Dark Matter eXperiment at the University of Washington in Seattle.

Pathogen transport

The US Centers for Disease Control and Prevention (CDC) on 11 July imposed a moratorium on the transfer

of biological materials from high-security biosafety labs, and temporarily closed its anthrax and flu labs. The move stems mostly from incidents this year in which CDC workers were potentially exposed to anthrax, and a harmless avian influenza virus shipped to an outside lab was contaminated with a highly pathogenic strain. See go.nature.com/lzqpwm for more.

PEOPLE

Energy nomination

US President Barack Obama nominated nuclear-policy adviser Elizabeth Sherwood-Randall on 9 July to be deputy secretary of the Department of Energy. Sherwood-Randall is currently the National Security Council's lead adviser on nuclear proliferation and defence issues. If confirmed by the Senate, her work at the energy department would include overseeing its programmes on nuclear weapons, energy research and energy production.

UK science minister

Greg Clark was announced as the United Kingdom's new minister for universities and science on 15 July, after previous incumbent David Willetts resigned in a cabinet reshuffle. Clark, who has a PhD in economics, adds the brief to his existing ministerial

COMING UP

20–25 JULY

The 20th International AIDS Conference takes place in Melbourne, Australia.

www.aids2014.org

responsibilities for cities, local growth and constitutional reform. Willetts had been in the role since 2010, during which time the science budget declined in real terms, but he was praised by leaders in the UK science community. See go.nature.com/itj9o4 for more.

POLICY

Health unit

The World Meteorological Organization and the World Health Organization announced last week that they have established a joint programme to address health risks arising from climate change and extreme weather. The climate and health office aims, for example, to forecast disease outbreaks and devise strategies to prepare for extreme heat, drought, floods and storms.

BUSINESS

Fast-track drug

A pharmaceutical firm has for the first time bought a candidate drug developed by the rare-diseases programme of the US National Center for Advancing Translational Sciences (NCATS). AesRx of Newton, Massachusetts, developed a drug called Aes-103 to treat sickle-cell anaemia in partnership with the centre. The purchase of AesRx and Aes-103 by Baxter, a drug company in Deerfield, Illinois, was announced on 9 July. NCATS was established in 2011 with the aim of getting experimental treatments to market more quickly.

➔ NATURE.COM

For daily news updates see:

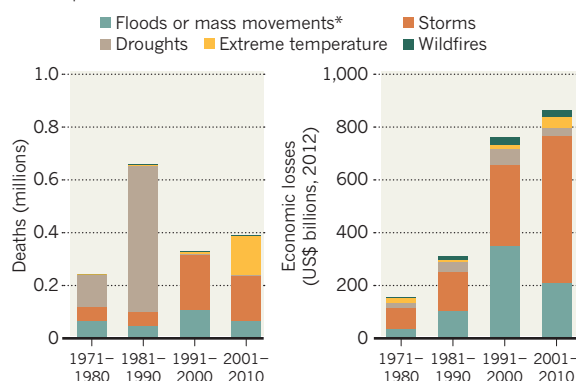
www.nature.com/news

TREND WATCH

Storms and floods accounted for 79% of the 8,835 weather, climate and water-related disasters between 1970 and 2012, and caused 54% of the deaths and 84% of the economic losses for these events, according to the World Meteorological Organization and the Centre for Research on the Epidemiology of Disasters in Belgium. The greatest number of disasters happened in 2001–10, although 1981–90 was the most deadly decade, owing to severe droughts and famines in Ethiopia, Sudan and Mozambique.

THE COSTS OF DISASTER

Weather, climate and water extremes caused 1.62 million deaths and US\$2.1 trillion in economic losses from 1971 to 2010.



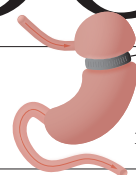
*Avalanches, landslides, subsidence, rockfalls

NEWS IN FOCUS

NATURAL DISASTERS Landslide prediction system debuts at world forum **p.272**

FUNDING Scripps merger failure highlights plight of private research institutes **p.274**

FUNDING CERN charity launch seeks donations for big physics **p.276**



WEIGHT LOSS How gastric surgery causes metabolic changes **p.282**

FOSSNI/PHOTOGRAPHIC SECTION, NAT/ROYAL NAVY



The Royal Navy's nuclear submarines could benefit from a navigation system based on quantum technology now under development.

PHYSICS

Quantum-hub finalists picked

UK government considers eight proposals for up to six research centres.

BY KATIA MOSKVITCH

The UK government has narrowed the list of teams competing for millions of pounds in quantum-technology funding down to the last eight. Pledged in December by UK chancellor George Osborne, the £270-million (US\$462-million) funding pot is primarily meant to establish up to six research hubs focusing on different applications of the rapidly advancing field.

The teams still in the running — led by

Imperial College London, University College London and the universities of Birmingham, Bristol, Glasgow, Lancaster, Oxford and York — will learn on 15 September whether their proposals have been accepted by the National Quantum Technologies Programme. “I think this is the biggest single investment in the emerging technologies that the UK government has ever made,” says Peter Knight, a physicist at Imperial College and former president of the Institute of Physics in London.

For decades, quantum physics seemed

too esoteric to have much practical use. But now, physicists see opportunities to take quantum research out of the lab and into real life. Technologies that harness the peculiar qualities of quantum mechanics seem poised to deliver breakthroughs in a wide range of applications.

“The pace of transition from theoretical concepts to realizable quantum technologies is astonishing,” says Nicola Wilkin, a theoretical physicist at the University of Birmingham. “Applications that will be rapidly realized ▶

► include unprecedented gravitational sensors to find oil and minerals, clocks that deliver next-generation navigation, and secure broadband communication.”

Of the £270 million, £190 million is new money; the rest will be drawn from the government’s roughly £1-billion annual allocation for science capital funding. Most of the money will be distributed through the Engineering and Physical Sciences Research Council (EPSRC).

“Each hub will cover a carefully selected theme, such as computing or communication, where quantum technologies can provide game-changing advances and benefit from enhanced collaboration between industry, academia and government,” then-science minister David Willetts told *Nature*.

Imperial College proposes a hub focused on manipulating the quantum states of ultracold atoms. An Imperial project already under way with funding from the UK Defence Science and Technology Laboratory aims to develop a quantum-based ultra-precise submarine positioning system for the Royal Navy. Submarines face a navigation challenge because they cannot contact positioning satellites without surfacing. “After six months of wandering around under the ocean, you could be way off where you think you are,” says Imperial physicist and project leader Edward Hinds. The system promises to be 1,000 times more accurate than today’s technology — with no need to surface. And because space is at a premium on submarines, the researchers also want to make the device smaller; their current model is 50 centimetres wide. The team hopes to have a narrower prototype available by 2016.

Other teams are focusing on different applications. The Lancaster University group, for example, aims to develop quantum sensors and metrology tools for use in health care and nuclear power, says Yuri Pashkin, director of Lancaster’s Quantum Technology Centre, which opened in May. And University of Oxford physicist Ian Walmsley says that his institution’s proposed hub would establish Britain as a global leader in quantum technologies for defence, communications, pharmaceuticals and finance by pursuing powerful computers, simulators, communications networks and sensors.

The goal of the investment is to secure the United Kingdom’s strong global position in quantum physics and keep British quantum physicists at home, says Rachel Bishop, theme leader for quantum technologies at the EPSRC in Swindon. “As a scientist, you want to work somewhere exciting where you can explore your ideas in a well-funded research environment — and that’s exactly what the government is doing in quantum.” ■



Plans are afoot to improve landslide monitoring for the endangered town of Zhangmu in Tibet.

NATURAL HAZARDS

Landslide risks rise up agenda

Forum on deadly natural phenomena discusses use of simulation and hazard-mapping technologies.

BY JANE QIU

The Tibetan town of Zhangmu is on edge — in an emotional and physical sense. Perched precariously on a mountainside, the growing trading and tourist centre lives under the constant threat of landslides, the result of a formidable combination of geological, climatic and developmental factors. The settlement, whose population reaches 40,000 in summer months, is built on the unstable debris of past landslides. As more buildings appear, the risk of a catastrophic collapse increases.

Many settlements across the globe face a similar predicament. With extreme weather events becoming more common, land resources dwindling and urban development spiralling, landslides “are increasing in frequency, scope and destructive capacity”, says Sálvano Briceño, chair of the scientific committee at Integrated Research on Disaster Risk, an international research programme headquartered in Beijing.

But the risks are being addressed. At the third World Landslide Forum in Beijing last month, researchers met to discuss ways to improve the monitoring, prevention and management of these lethal phenomena. Presentations included technologies for mapping hazards and providing early warnings, as well as computer models that simulate the effects of rainwater and earthquakes. “With the projected increase in extreme rainfall, communities in landslide-prone regions will be more vulnerable,” said Rex Baum, a geologist with the US Geological Survey in Golden, Colorado.

Slope failures are the biggest landslide threat. These occur when a chunk of slope becomes detached from a hillside. As the material descends, shearing forces increase the pressure of water in the gaps between soil and rock particles (the pore-water pressure), causing clumps of slope materials to collapse. This process, called liquefaction, can be a result of rainfall-induced increases in water volume or seismic



PUREU ZHAXI/XINHUA PRESS/CORBIS

waves, and greatly accelerates the landslide because the water acts as a lubricant.

Catastrophic slides are frequent. In 2010, heavy rains in Zhouqu in northwestern China unleashed a torrent of mud and rocks that engulfed 550 houses and killed nearly 1,800 people. And in May this year, a rain-drenched scarp in northeastern Afghanistan gave way, sweeping away the village of Ab Barak and killing more than 2,000 inhabitants.

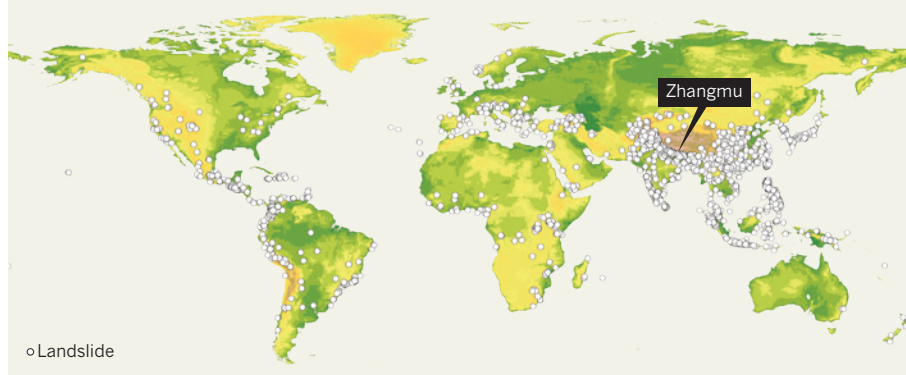
Developing countries are worst hit (see 'Danger zones'). A study by geologist Dave Petley at Durham University, UK, shows that, of the 32,322 landslide fatalities between 2004 and 2010, most occurred in Asia, especially in the Himalayas and China (D. Petley *Geology* **40**, 927–930; 2012).

But advances in remote-sensing technologies are making hazards easier to detect. Satellite and airborne laser and radar instruments, such as LiDAR and InSAR, can be used to monitor ground movements, enabling accurate mapping of potential landslide sites.

"We are getting pretty good at spotting areas susceptible to landslides," says Baum. "But we still can't quite predict, if a slope fails, how big it

DANGER ZONES

Between 2004 and 2010 there were 2,620 fatal landslides worldwide, causing 32,322 deaths. Most occurred in Asia and other developing parts of the world. China and the Himalayan region were particularly badly affected.



will be or how far it's going to go." The landslide that struck in Washington state on 22 March, killing 41 people, took many by surprise. "We didn't really expect that a slope coming off a block that was only 200 metres high could have flowed over one kilometre," Baum says.

A big unknown is how rainfall, which triggers two-thirds of landslides, can change groundwater dynamics and the strength of soil and rock particles, says Kyoji Sassa, a geologist at Kyoto University in Japan. At the forum, his team presented a lab-based landslide simulator that tests how pore-water pressure and the strength of slope materials change with increasing rain. By feeding the data into a computer model designed to reproduce both the initiation and movement of a slide — a first for a landslide model — they have been able to replicate past events.

In a US\$5-million project funded by the Japanese government, Sassa and his colleagues are testing the approach on a notoriously unstable slope in southern Vietnam, where annual precipitation is more than 4,000 millimetres. They will combine rainfall records and weather forecasts to see if the simulator and model can predict how the slope will react to further rain. The ultimate goal, says Sassa, is "to develop a model that could be applied in all monsoonal regions".

In the meantime, Zhangmu, which is prone to earthquakes and heavy rainfall, needs a contingency plan. A survey led by Wei Fang-qiang, deputy director of the Chinese Academy of Sciences' Institute of Mountain Hazards and Environment in Chengdu, found that the 49–78-metre-deep layer of previous landslide

debris below the city is already moving, albeit slowly. It identified 21 potentially dangerous sites, some of which could produce several million cubic metres of debris (the Washington slide generated about 7.6 million cubic metres).

Last month, the Chinese government approved a \$483-million project to improve monitoring in the Zhangmu region. Engineers will install sensors to determine pore-water pressure and implement measures to stabilize slopes, drain rainwater and block debris flow.

Critics warn that many governments tend to invest much more in disaster mitigation and relief than in reducing exposure to hazards.

"What kills people are not natural phenomena, but poorly built or wrongly located houses."

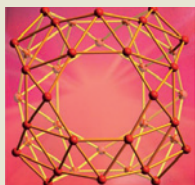
"Many mountainous regions are being developed rapidly without proper planning or risk assessment," says Briceño. "What kills people are not natural phenomena, but poorly built

or wrongly located houses."

The second phase of the United Nations' Hyogo Framework for Action, a ten-year plan aimed at reducing the impact of natural disasters, including landslides, should help to address such problems, he adds. Its latest incarnation, which is expected to tackle the challenges of extreme climate events and land-use changes, is due to be adopted next March. "Risk reduction is the key," says Briceño. "It should go hand in hand with climate-change adaptation and sustainable development." ■



TOP STORY



First 'buckyballs' made of boron instead of carbon
go.nature.com/tyzalq

MORE NEWS

- Starvation in pregnant mice leaves epigenetic marks in offspring go.nature.com/ajhnyw
- Oxygen fluctuations drove early evolution go.nature.com/89avhj
- Chimp intelligence is hereditary go.nature.com/wlsnaj

NATURE PODCAST



Why cancer patients waste away; compressing diamond; and the Longitude exhibition nature.com/nature/podcast



EDUARDO CONTRERAS/SAN DIEGO UNION-TRIBUNE/ZUMA PRESS

Michael Marletta's plans for the Scripps Research Institute have prompted a declaration of no confidence in their president from faculty members.

FUNDING

Scripps merger fiasco highlights US funding woes

Other independent biomedical research institutions have turned to private benefactors.

BY ERIKA CHECK HAYDEN

Faced with vehement opposition from faculty members, a prestigious independent research institute in California has abandoned a planned merger with a nearby university just weeks after the proposed deal became public. The brief affair between the La Jolla-based Scripps Research Institute and the University of Southern California (USC) in Los Angeles exposes a growing and perhaps insurmountable rift between Scripps faculty members and the institute's president, Michael Marletta, who forged the deal to help close a projected US\$21-million operating deficit. "I think there is a misunderstanding of what this institute is about," says Scripps ophthalmologist and molecular biologist Martin Friedlander, echoing the sentiments of many.

The crisis also dramatically illustrates problems faced by many independent US research institutes, which tend to be more vulnerable

than universities to the vicissitudes of relying on funding from the National Institutes of Health (NIH). Other independent institutes in La Jolla have partially compensated for the decline by intensifying their private fundraising efforts. Scripps, which conducts basic biomedical research, draws 86% of its operating budget from the NIH, according to a 2013 report by the ratings agency Fitch, at a time when the agency's overall funding is declining when adjusted for inflation. At the same time, financial support from pharmaceutical companies, a second major source of funding for Scripps, has all but dried up.

"They have become dependent on two things that disappeared at the same time," says Kristiina Vuori, president of the nearby Sanford-Burnham Medical Research Institute.

Scripps has also expanded greatly in the past few decades. Richard Lerner, who led the lab from 1987 to 2012, more than quadrupled the institute's staff during his tenure, and in 2003 opened a second campus in Jupiter, Florida.

The evaporation of pharmaceutical support has been especially painful. Scripps had long benefited from deals in which drug companies provided financial support for basic research in exchange for first rights to intellectual property arising from any discoveries. But those arrangements have fallen out of favour in recent years. Scripps' last big deal, with New York city-based Pfizer, brought the lab \$100 million over five years. It ended in December 2011, a month before Marletta became president.

Marletta's ill-fated attempt to reduce Scripps' deficit put a price of \$600 million on the institute, to be paid by USC over 40 years. But in what seems to have been a tactical error, he told faculty members about his negotiations with USC only in June, when the deal was all but signed. They immediately objected, asking the Scripps board to keep the lab independent. In early July, most of the faculty members, including all of the institute's department chairs and its dean, signed a letter to the board declaring that they had no

confidence in Marletta's leadership.

"In a way, it was like going out of the family," says Evan Snyder, a stem-cell researcher at Sanford-Burnham who has close collaborations with Scripps researchers, explaining the faculty members' vehement opposition to the deal. They distrusted USC, an academic monolith seen as less prestigious and more bureaucratic than Scripps. "The personalities are different between Scripps and USC," says Snyder, adding that a merger offer might have been better received had it come from one of the other independent institutes clustered on the cliffs of La Jolla.

Scripps faculty members also felt that the deal sold them short. In interviews, they noted Scripps' coveted ocean-front location: La Jolla is one of the priciest zip codes in the United States. The \$15-million annual payments over 40 years offered by USC would be the equivalent of a \$250-million mortgage, they say. That would not even cover one year of the institute's operating expenses, which were \$400 million in 2013.

"It didn't make a lot of sense financially," Friedlander says. "You can't ignore a \$20-million deficit, but there are many other creative ways of addressing the financial shortfall. We certainly do not have our backs against the wall."

Although Scripps researchers did not

necessarily object to the idea of some relationship with USC, they worried about how much independence the institute would retain in the deal — for instance, whether it would still run its own graduate training programme. "If Scripps was fairly valued and the financial terms gave Scripps financial security, the deal would have been viewed differently, and people would be prepared to step up and talk about all the other issues," one faculty member says.

USC officials responded to the merger's demise with a brief statement from provost Elizabeth Garrett emphasizing the university's willingness to collaborate with Scripps researchers and its commitment to "exceptional biomedical research programs that produce meaningful scientific discoveries and benefit patients throughout the world".

What will happen next is not clear. Marletta and Richard Gephardt, chair of the Scripps board of trustees, announced on 9 July that the deal was dead and that a committee of administrators, faculty and board members is "reviewing a broad range of thoughtful

"You can't ignore a \$20-million deficit, but there are many other creative ways of addressing the financial shortfall."

alternatives to choose the best path forward for the institution".

Faced with similar problems, Scripps' neighbouring institutes have greatly ramped up their efforts to raise money through philanthropy. Sanford-Burnham, which has long benefited from the support of two businessmen, received an anonymous \$275-million donation in January. The Salk Institute for Biological Studies, also in La Jolla, has raised \$275 million in a campaign that it hopes will reach \$300 million by next year.

Marletta has said that he is seeking more donations for Scripps, but is disadvantaged by being a relatively recent arrival; he was chair of the chemistry department at the University of California, Berkeley, until 2010.

"Philanthropy is about long-term relationships with your donors; it's not something where you just turn the spigot and say, 'OK, we'll go out and raise a billion dollars,'" says Salk president William Brody, who initiated his institute's fund-raising campaign soon after arriving in 2009.

Still, Brody and other observers say that Scripps should be able to find a way out of its current dilemma that does not involve dissolution or losing its independence.

"If they can stick to their knitting and stay the course, they will be successful," Brody predicts. ■ [SEE EDITORIAL P.263](#)

Charity begins at CERN

Particle-physics lab sets up fund for 'extras' as other big institutes mull similar move.

BY ELIZABETH GIBNEY

There is a mantra in the fund-raising world: big donors like to support big ideas. And ideas do not come much larger than at CERN, Europe's particle-physics laboratory near Geneva in Switzerland. Now the organization — which uses its particle smasher to probe the fundamental structure of the Universe — has registered a charitable foundation to raise funds for its educational, technology-transfer and arts activities.

CERN is not the only big institution to go after donations to fund projects that fall outside the core research remit. The trend is on the rise among large European research organizations. The European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany, is shifting its fund-raising focus from industry sponsorship to private donations. And ITER, the international nuclear-fusion experiment being built in Cadarache, France, is devising a way to deal with the offers of donations that it already receives. What nobody yet knows is the fruit these efforts will bear — whether individuals really want to donate heftily to scientific charities that are not focused on medical solutions.

For CERN, there is no better time to form a charitable foundation, says Matteo Castoldi, head of its development office. CERN's Large Hadron Collider, and the discovery of the Higgs boson, has "captured the public imagination" as much as the Apollo missions did in the 1960s, he says. The organization is already taking advantage of this, "but there is much more we could do, and that's where the foundation comes in".

Registered in Switzerland last month, the CERN & Society foundation is designed to put CERN's fund-raising efforts on a firmer footing: although the lab has accepted donations in the past, charitable status means that donors can now pledge tax-deductible gifts. Organizers hope that this will encourage more — and larger — donations.

CERN director-general Rolf-Dieter Heuer stresses that such funding will not replace the institute's core budget, paid for by member states. Instead, the proceeds are aimed at activities that this funding cannot stretch to: school projects, the development of medical spin-offs such as proton therapy (the use of proton beams to kill cancer cells), and meeting the huge demand for general-interest and science-related visits. But if a donor has an explicit desire for



Visitors to CERN currently number roughly 90,000 a year, but with more funds the lab could host about 300,000.

their gift to go towards research, CERN would consider this, adds Heuer.

Continental Europe has been slow to embrace professional fund-raising and a wider culture of philanthropy: it lags behind the United Kingdom by about 20 years and the United States by 50 years, says Johannes Ruzicka, managing director of the fund-raising consultancy Brakeley in Munich, Germany, which is advising EMBL. More research institutions are thinking about this kind of fund-raising, he adds, but few actually make the leap, owing to the significant investment and administrative hassle that goes with setting up a foundation.

European universities have been bolder than research institutions, and have been trying to emulate the fund-raising abilities of their US counterparts for some time, says Kate Hunter, executive director of Europe's arm of the Council for Advancement and Support of Education, based in London. "There's been a massive trend over the last decade to reinforce that universities

are charitable entities in their own right, and that they are a legitimate cause to support," she says. "So I do think it's an interesting development if pure science research institutions see that they can do that too."

There are reasons for the reticence of institutions. Some facilities that are funded by several countries fear that raising large amounts of money through philanthropy could encourage governments to cut their contributions, says Ruzicka. But others believe that state-funding agreements have a finite lifespan and so need to be backed up by other funding sources, he adds.

ITER's goal — to build an experimental fusion reactor that will serve as a stepping stone towards harnessing effectively limitless energy — already makes it an attractive candidate for philanthropists. The facility is creating its charity framework directly in response to people asking to contribute, says an ITER spokeswoman. The cash will go towards educational activities, internships, exhibitions and conference travel costs, although ITER is not yet authorized to accept tax-deductible donations.

CERN has no history of professional fund-raising, and Castoldi acknowledges that progress will be slow. It also remains to be seen to what extent particle physics will appeal to philanthropists. Hunter is optimistic. "Places like CERN and other research institutions are doing amazing things that will ultimately deliver public benefit, so if those organizations can make that case, that can be quite attractive for donors," she says.

Heuer says that CERN is "completely open" to offers of any size, and Castoldi hopes that it will raise 25 million Swiss francs (US\$28 million) in the next five years. Individuals, trusts and companies can donate, and contributors will be recognized in various ways.

Those who make substantial gifts could even have a facility at CERN named after them, says Heuer — but not, he adds, any particles the laboratory might discover. "That is science," he says. "We don't touch that." ■

CORRECTION

The timeline in the News Feature 'Hope on the line' (*Nature* **511**, 19–21; 2014) wrongly identified Alan Trounson as the first president of the California Institute for Regenerative Medicine. He was, in fact, the second — he succeeded Zach Hall.

DENIS BALBOUSE/REUTERS/CORBIS



WAVE OF THE FUTURE

BY ALEXANDRA WITZE

After two decades and more than half a billion dollars, LIGO, the world's largest gravitational-wave observatory, is on the verge of a detection. Maybe.

In the Louisiana swamps just east of Baton Rouge, the daily hunt for gravitational waves cannot really get started until well after noon.

Mornings are a lost cause, thanks to the sonic chaos from traffic rumbling along the nearby interstate highway, trains roaring past and loggers occasionally unleashing their chainsaws on plantations of pine trees.

Even now, at 6 p.m. on a weekday evening in May, Ryan de Rosa is gazing with resignation at a set of computer monitors in the control room of the Laser Interferometer Gravitational-Wave Observatory (LIGO). The displays are starting to stabilize, but they still show the myriad jolts — imperceptible to humans — that are rocking the ground. The traces,

Technicians work on part of the LIGO gravitational-wave detector in Livingston, Louisiana.

generated by distant earthquakes, traffic and even waves breaking on the coast of the Gulf of Mexico more than 100 kilometres away, look like jagged mountain peaks.

De Rosa, a physicist at Louisiana State University in Baton Rouge, knows he has a long night ahead of him. He and half a dozen other scientists and engineers are trying to achieve 'full lock' on a major upgrade to the detector — to gain complete control over the infrared laser beams that race up and down two 4-kilometre tunnels at the heart of the facility. By precisely controlling the path of the lasers and measuring their journey in exquisite detail, the LIGO team hopes to observe the distinctive oscillations produced by a passing gravitational wave: a subtle ripple in space-time predicted nearly a century ago by Albert Einstein, but never observed directly.

Within weeks of this May evening, de Rosa and his colleagues will finally achieve full lock. A team working on an identical LIGO detector at the Hanford nuclear complex in Washington state should get there within months. If all goes well, the dual devices — which have together cost some US\$620 million — could resume taking data next year. They will be the most sensitive of several gravitational-wave detectors around the world that are racing to be the first to claim a discovery.

The anticipation and competition are intense. Finding direct evidence of gravitational waves would launch a new era of astronomy. Spotting not just one gravitational-wave source, but eventually dozens and then thousands, astronomers say, will give them new ways to watch black holes collide, stars annihilate themselves and space-time shimmy. Gravitational waves would thus open an entirely new window onto a dynamic, ever-changing universe.

There is just one problem. The first incarnation of LIGO hunted the waves for nearly a decade — and found none. Now, with the major upgrade, the project faces the hard reality of having to finally deliver on its promises.

EVERYWHERE AND NOWHERE

In theory, Earth should be awash in gravitational waves. They are thought to come from any cosmic event that disturbs the fabric of space and time with sufficient force, in much the same way that seismic waves radiate from an earthquake. A dying star that explodes as a supernova should produce a tsunami of gravitational waves. More-rhythmic waves might come from the rotation of a dense object that is not quite perfectly symmetrical — say, a furiously spinning neutron star with a small bulge in its side. Another source might be a pair of black holes or neutron stars that whirl around one another, gradually drawing closer until they collide in a final, catastrophic merger.

That last example is not hypothetical: in 1974, using the Arecibo radio telescope in Puerto Rico, physicist Joseph Taylor at the University of Massachusetts Amherst and his then-graduate student Russell Hulse discovered just such a neutron-star binary. Over the next few years, Taylor and Hulse watched the timing of radio flashes from one of the spinning stars change ever so slightly as the pair spiralled closer. The shifts matched Einstein's prediction of how gravitational waves would carry energy away from an imminent stellar smash (R. A. Hulse and J. H. Taylor *Astrophys. J.* **195**, L51–L53; 1975). It was the first indirect detection of gravitational waves, and it netted Hulse and Taylor the 1993 Nobel Prize in Physics.

The first attempt to observe gravitational waves directly had come in the early 1960s, when Joseph Weber of the University of Maryland in College Park tried unsuccessfully to observe vibrations caused by gravitational waves passing through an aluminium cylinder. Then, in the late 1960s, physicist Rainer Weiss proposed using lasers rather than a metal bar. The concept involves splitting a laser beam into two using an elaborate maze of mirrors, and sending them down two tunnels that are set at right angles to one another, and back again. The set-up takes advantage of the polarized nature of gravitational waves: when they pass through an object — in this case,

the tunnels — they cause it to expand ever so slightly in one direction and contract in the perpendicular direction. Weiss, of the Massachusetts Institute of Technology in Cambridge, suggested it would be possible to detect that kind of warping by re-combining the separated laser beams and using interferometry to look for tiny shifts in the way they interact (see 'To catch a wave').

In 1992, after decades of planning, replanning and prototyping, the US National Science Foundation (NSF) committed to spending \$272 million (\$420 million in 2008 dollars) on building such an interferometer, now called LIGO. The plan called for two identical detectors separated by thousands of kilometres, so that the observatory could cross-check its own results: sites in Washington and Livingston, Louisiana, were chosen.

"WITH ADVANCED LIGO, DETECTIONS WOULD BE PROBABLE."

What the plan did not call for was a gravitational-wave discovery — at least not any time soon. "We had this careful choice of words and a story about what we were going to do," says Barry Barish, a physicist at the California Institute of Technology (Caltech) in Pasadena, who helped to make the case to the NSF and became LIGO's principal investigator in 1994. First there would be the initial LIGO, which would develop and demonstrate the technology, with any discovery coming as a bonus. And then would come a second stage — Advanced LIGO, which would require a separate go-ahead from the NSF, and would increase the sensitivity by an order of magnitude. "We said that with initial LIGO, detections would be possible," says Barish, "and with Advanced LIGO, detections would be probable."

The problem was that estimates of what LIGO would see were still very uncertain. "When we initially proposed LIGO, the only sources that we were really contemplating were supernovae," says Weiss. "We thought we would see something like one a year, maybe even ten a year." But then improved computer simulations radically downsized the amount of gravitational-wave energy that would be expected from such explosions. A supernova would have to go off very close to Earth for LIGO to see anything from it.

Other calculations cut back on how often LIGO would be expected to see gravitational waves from lone wobbly neutron stars. "There was an optimism about sources that turned out not to have been justified," says Cole Miller, a theoretical astrophysicist at the University of Maryland who chaired LIGO's external science advisory panel until last year.

But by the time the observatory got the go-ahead, the LIGO scientists were growing more optimistic about pairs of neutron stars. They realized that when these stars collided they would send out a clean, easily detectable gravitational-wave signal right in the frequency range where LIGO was most sensitive. Even at its relatively low initial sensitivity, the observatory could have detected two neutron stars merging anywhere within 20 megaparsecs (65 million light years) of Earth. Yet it was still a long shot, says David Reitze, executive director of the LIGO Laboratory, who is based at Caltech: "We would have had to have gotten lucky."

They were not. During the first phase of LIGO, from 2002 to 2010, Hanford and Livingston saw nothing. Still, the NSF was satisfied enough with LIGO's progress that it allocated another \$205 million for Advanced LIGO in 2008.

The upgrade will slowly increase the sensitivity of the detectors by a factor of ten, so that Advanced LIGO will be able to see neutron-star mergers not at 20 megaparsecs, but at 150 or even 200. That will multiply the volume of space that LIGO can search by 1,000, and will vastly improve the chance that the detector will spot one of the rare events that produce a gravitational wave.

NATURE.COM
Read more about
LIGO:
go.nature.com/mm4ba8

Current best estimates of neutron-star merger rates suggest that with any luck — assuming that neutron stars don't collide at the absolute lowest end of the probability range, and do go off within the search volume during the observation period — Advanced LIGO will see several of them per year once it reaches its design sensitivity. “The real question is not whether we're going to detect gravitational waves, but will they come frequently or will they come rather rarely,” says Stanley Whitcomb, a longtime LIGO physicist at Caltech who serves as the project's chief scientist.

NOISY NEIGHBOURS

But first, the LIGO team has to finish building the advanced system. In 2011, engineers began ripping components out of the tunnels at the Livingston and Hanford sites to replace them with much more elaborate versions. LIGO's performance is determined by how accurately it can measure distortions created by a passing gravitational wave in the length of the interferometer's 4-kilometre arms. In its initial configuration, the observatory was able to measure those distortions to about one part in 10^{21} — equivalent to a shift of about one-thousandth the diameter of a proton. To improve the sensitivity by a factor of ten, Advanced LIGO's designers have made a number of major changes, starting with better ways to isolate the machine from random ground-shaking.

Seismic noise is a problem particularly at Livingston, where the detector sits just a few kilometres from a major interstate highway and a railway line. Surveys as far back as 1988 had warned about noise there, but the problem did not seem insurmountable. And Louisiana senator Bennett Johnson (Democrat), who was on the panel that appropriated money for the NSF, helped to push the project through. Livingston did have some practical advantages, including few earthquakes, lots of flat land and proximity to an established group of gravitational physicists at Louisiana State University. Planners thought that they could compensate for the noise with a range of devices to dampen ground motion.

They couldn't, at least at first. When trains blasted by during the earliest science runs, the interferometer shook so much that it was knocked offline. Even worse was the local logging. Brian O'Reilly, a senior scientist at the Livingston lab, calls it “the constant bane of our existence”. He waves his hand in frustration out of the window of his office, towards a plot of land just off the LIGO property that was clear-cut during early detector operations. “It wasn't like we could say, ‘Please stop your multimillion-dollar industrial effort so we can detect gravitational waves.’”

“MY NIGHTMARE IS THAT IT
HAPPENS BEFORE WE TURN IT ON.”

But the logging is a problem only occasionally, and over time LIGO engineers have fine-tuned the system to withstand passing trains.

Looking like a proud parent, O'Reilly uses a scale model of Advanced LIGO to point out a host of obsessive changes made to the noise-isolation system. In each of the arms, the mirrors that reflect the laser beam hang from glass cylinders, which in turn hang from metal plates that hang from yet other plates. Each layer of suspension provides another opportunity to dampen unwanted vibrations. Amid all the glass and metal, triangular steel blades serve as extra protective isolators, delicately balancing the weight of three-quarters of a tonne of engineering equipment.

Advanced LIGO also incorporates more-powerful lasers, plus a set of recycling cavities that essentially trick the detector into thinking that there are more photons in it than there are, boosting sensitivity. (There is an upper limit to how much light can actually be pumped into LIGO,

because the more photons there are, the more they contribute to a white-noise-like effect at high frequencies that ruins the signal.)

Although the system looks perfect in the scale model, the actual project has run into construction difficulties. At Hanford, the material that coats the hanging glass mirrors showed some unexpected deterioration, so two of them are being replaced. At Livingston, mud-dauber wasps made nests in the insulation surrounding the beam tube, where their chlorine-rich excretions — which in part came from eating poisonous black-widow spiders — caused a leak in the vacuum system. The leak has been fixed and the wasps cleared out.

Even so, as of the night of 29–30 June the Livingston detector has managed to achieve full lock for more than two hours at a time, pulling off an official milestone months earlier than expected. If commissioning continues to go relatively smoothly, plans call for the first Advanced LIGO observing run to start in late 2015. A second run, with a decent shot of finding a gravitational wave, would occur in the winter of 2016–17. (Weiss likes to point out that a 2016 discovery would be a nice 100th-anniversary commemoration of Einstein's paper describing gravitational waves.) By the third science run, planned for 2017–18, the machine should be getting sensitive enough to almost certainly nail a detection, says Reitze.

This schedule, however, depends heavily on how quickly engineers can commission both interferometers. The team has decided to focus its energies on commissioning the detector at the relatively low frequencies where signals from binary neutron stars are thought to lurk. They will not worry so much about improving LIGO's performance at high frequencies, to snag other types of signals such as colliding black holes, unless they have their first gravitational waves in the bag.

GLOBAL COMPETITION

There are other groups out there seeking gravitational waves, and they just might beat LIGO to the punch. Like light, gravitational waves come in a huge variety of wavelengths — and just as radio telescopes and X-ray telescopes reveal different phenomena, so too should gravitational-wave detectors working at different wavelength ranges. “Each one of these experiments is doing something exciting,” says David Shoemaker, a physicist at MIT and head of Advanced LIGO.

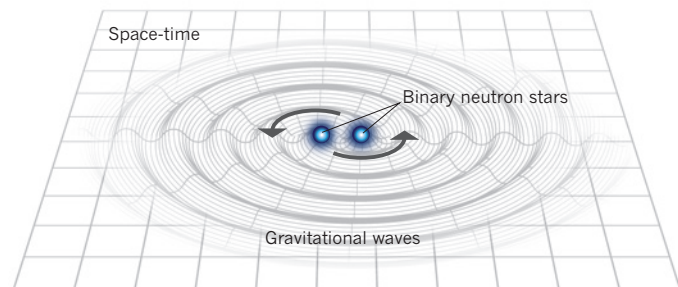
In March this year, there was a burst of gravitational-wave excitement about a report that the BICEP2 telescope at the South Pole had detected primordial gravitational waves left over from cosmic inflation that occurred moments after the Big Bang (see *Nature* **507**, 281–283; 2014). The wavelengths of these disturbances essentially span the entire Universe, far outside the wavelength range that LIGO can see. The BICEP2 team initially reported a strong signal, but when the scientists published their findings in June (R. A. R. Ade *et al.* *Phys. Rev. Lett.* **112**, 241101; 2014), they admitted that they could not rule out the possibility that the gravitational-wave ‘signal’ was just an artefact of galactic dust (see go.nature.com/lruz8e).

A very different kind of hunt is under way by a North American–European–Australian collaboration of astronomers who have been monitoring about 70 pulsars: rapidly spinning neutron stars that emit signals at incredibly precise intervals. The members of the International Pulsar Timing Array (IPTA) hope to detect a passing gravitational wave by the way it affects the timing of the pulses. They would have to be very lucky to see one before Advanced LIGO does, says IPTA co-leader Scott Ransom, an astronomer at the University of Virginia in Charlottesville. But even so, he says, “I always tease the LIGO people that here comes the dark horse”.

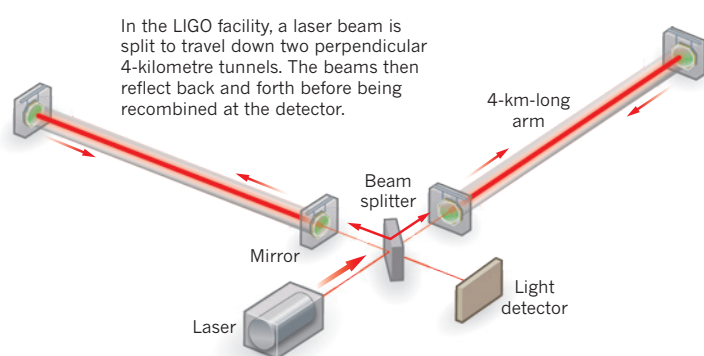
The gravitational waves found through pulsar timing would also be very different beasts from the ones LIGO is seeking. They would come from sources such as colliding supermassive black holes, whose huge mass would make their coalescence frequency much too low for an interferometer like LIGO to see. Nevertheless, says Joseph Giaime, head of the Livingston observatory, any direct detection will invigorate the field. “You can only go so many decades without detecting anything before some people start to think there's some quackery involved.”

TO CATCH A WAVE

The Laser Interferometer Gravitational-Wave Observatory (LIGO) is trying to detect ripples in the fabric of space-time predicted by Einstein's general theory of relativity.

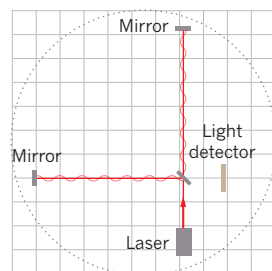


These ripples are thought to be produced whenever moving masses distort the space-time around them. A particularly powerful source might be a pair of neutron stars or black holes whirling around one another in a close orbit.

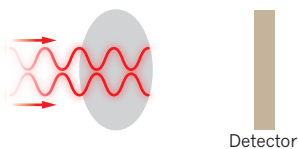
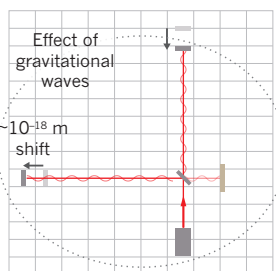


In the LIGO facility, a laser beam is split to travel down two perpendicular 4-kilometre tunnels. The beams then reflect back and forth before being recombined at the detector.

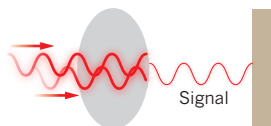
Normally, the two light beams travel paths of identical lengths, so that they cancel each other out when they recombine at the detector.



When a gravitational wave passes LIGO, the tunnels deform slightly and the distance travelled by each beam changes so that they no longer cancel out. This produces a measurable signal at the detector.



Detector



Signal

LIGO is partnering with similar observatories around the world so that any signal can be independently verified, and its source triangulated.



The closest thing Advanced LIGO has to a competitor is also its closest ally. Virgo in Cascina, Italy, is like LIGO's little sister: a laser interferometer with 3-kilometre arms, it can reach only about three-quarters of LIGO's sensitivity.

Virgo hunts the same sources as LIGO, focusing mainly on colliding neutron stars. It began running in 2007 and has spotted no gravitational waves so far. But it, too, is in the middle of a major upgrade, currently scheduled to come online about a year after Advanced LIGO. Scientists from the two detectors share their data and collaborate closely; combining signals makes the analysis more robust, says Giovanni Losurdo, project leader for Advanced Virgo at the National Institute for Nuclear Physics in Florence, Italy. Crucially, having another interferometer on a different continent will help astronomers to accurately locate the source of any gravitational-wave signals.

While Virgo and LIGO are offline for upgrades, a third machine monitors the skies. GEO600 — an interferometer in Hanover, Germany, with two 600-metre-long arms — is much less sensitive than its bigger peers, but will be better than nothing if a big gravitational-wave-producing event does occur. This became clear in late May, when NASA's space telescope Swift reported a high-energy outburst in the nearby Andromeda galaxy. It turned out to be a false alarm, but had it been a real star explosion so close, both LIGO and Virgo would have missed the chance at a once-in-a-lifetime event. "My nightmare is that it happens before we turn on," says Gabriela Gonzalez, a physicist at Louisiana State University and spokesperson for the LIGO scientific collaboration.

Japanese scientists are building yet another interferometer: the Kamioka Gravitational Wave Detector (KAGRA), which will be buried deep in a mine and could be operational as early as 2018. And in Europe, researchers are dreaming of the Einstein Telescope, with three 10-kilometre arms buried in a triangle. But with a pricetag of at least €1 billion (US\$1.4 billion), the Einstein Telescope remains only a hope for now. Similarly, the European Space Agency has pushed back the proposed launch of a space-based gravitational-wave hunter, the Laser Interferometer Space Antenna (LISA), to 2034.

Even as project leaders try to get Advanced LIGO up and running, they are also pushing to place a third detector in India, where it would allow astronomers to pinpoint the source of gravitational waves even more accurately. LIGO engineers have already built a set of components, and are storing them at Hanford. They are waiting for India's new government to select a site and approve funding, but depending on when that happens, LIGO India could be operational by 2022 for a total cost of roughly \$350 million.

Back in the United States, Advanced LIGO has money to run until October 2018. If it has not reached its full design sensitivity by then, it will be almost certain to get operational funding from the NSF to keep trying for another five years, scientists say. Further upgrades to reduce noise at high frequencies could improve its sensitivity even more.

But although most physicists are optimistic that Advanced LIGO will eventually make a discovery, there is no guarantee. "If we get to the design sensitivity and make no detections, then there are a lot of things that will have to go back to the drawing board theoretically," says Barish. "If we fail, we're not expecting that the NSF will help bail it out somehow."

For now, the field's future rests in the hands of de Rosa and his colleagues. He frowns, perplexed, at a glowing screen in the Livingston control room. Something is still not quite right with how the light is bouncing off one particular mirror in the machine. But it is dinner time. He rounds up the others in the room, and they head for a Mexican restaurant for a short break.

As they pull out of the car park, a series of spikes appear on the LIGO monitors. The ultrasensitive detectors have picked up the rumbling from the researchers' cars, heading off into the night. ■

Alexandra Witze is a correspondent for Nature based in Boulder, Colorado.

A gut-wrenching question

Gastric-bypass surgery can curb obesity as well as diabetes and a slew of other problems. Researchers are now trying to find out how it works.

BY VIRGINIA HUGHES

Every week, about 20 people visit the University of Pittsburgh Medical Center in Pennsylvania to be evaluated for weight-loss surgery. They tell a nurse their medical history and have a routine physical examination. Then they sit down with a surgeon to discuss their options.

Anita Courcoulas, head of minimally invasive bariatric and general surgery at the centre, has had thousands of these conversations in the past 25 years. During that time, the information she shares with her patients has changed dramatically. Thanks to clinical trials, she can now tell them with some confidence that surgery not only spurs remarkable weight loss in most people, but also significantly lowers the risk of heart attack, stroke, cancer and death. And with the most popular procedure — Roux-en-Y gastric bypass, which shrinks the stomach to the size of an egg — up to 60% of patients with diabetes go into remission for at least several years after the operation¹.

There are drawbacks for her to discuss, too: the cost (around US\$25,000); the small risk of surgical complications (on a par with that of gall-bladder removal); and the chance of developing nutritional deficiencies or an intolerance to certain foods. But perhaps the toughest issue for patients is the uncertainty.

Surgery does not work for everybody, and weight loss can be transient.

Doctors are not sure why gastric bypass and similar procedures curb diabetes and other diseases. The conventional view has been that the benefits stem mostly from the weight that patients shed — typically one-quarter of their body mass¹. But in the 1980s, some patients were found to show rapid changes in their metabolism after surgery, suggesting that other factors are at play. Now, a slew of high-profile animal studies is identifying potential mechanisms in how the gut adapts to its strange new configuration: with sweeping changes in bacterial populations, bile acids, hormone secretions and tissue growth. The hope is that more research on what happens after bariatric surgery will enable physicians to identify who will respond best — and even lead to ways of altering metabolism without resorting to the knife.

HUNGER STRIKE

Bariatric surgery debuted in Sweden in 1952, when surgeon Viktor Henrikson removed a 105-centimetre stretch of a woman's small intestine. The procedure did not help the woman to lose weight, but it did treat her constipation and boost her metabolism. According to Henrikson's case report², she was "content, subjectively felt healthier and more energetic".

Over the next two decades, surgeons in the United States refined the procedure. They cut the small intestine near each end,

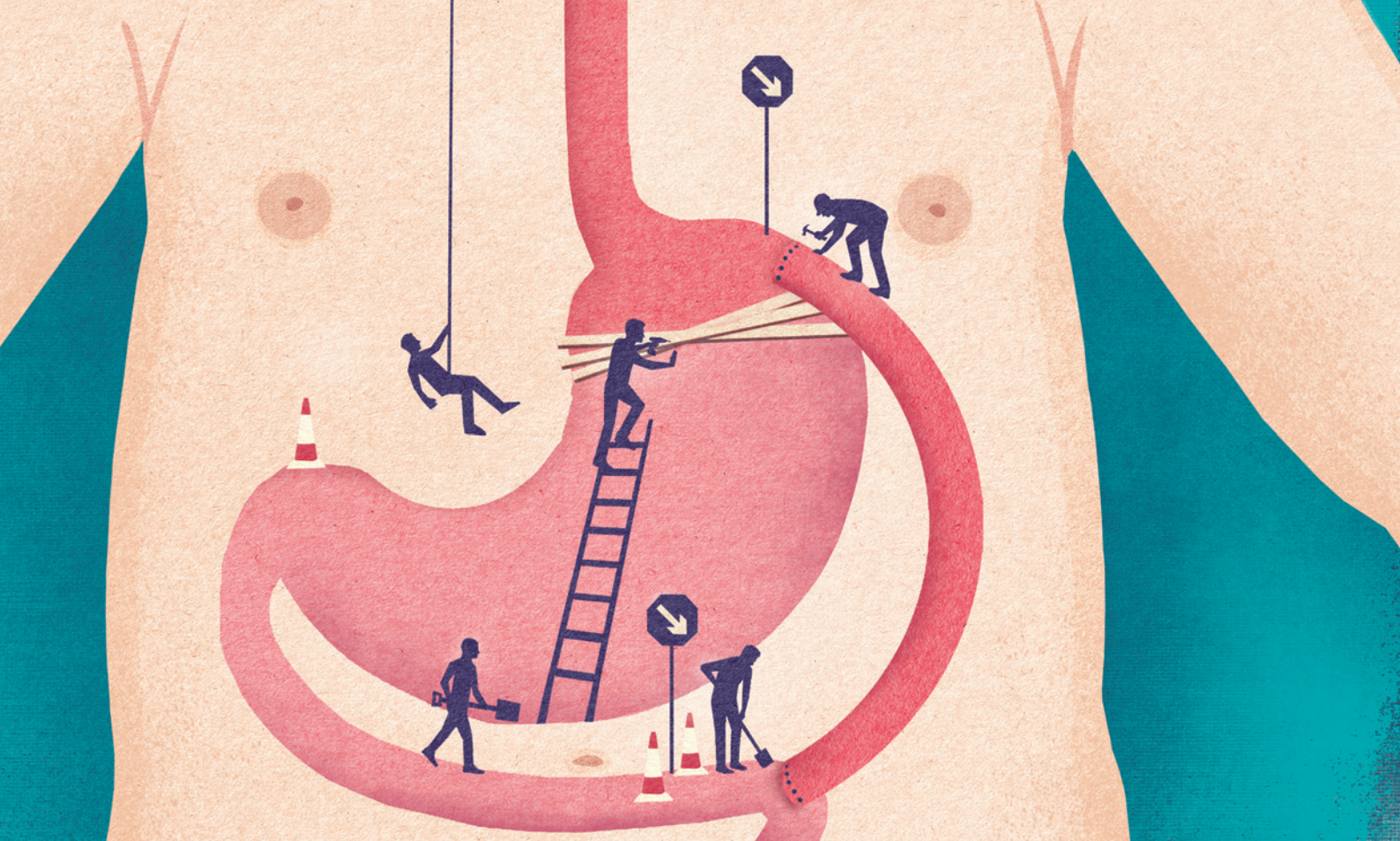
then rejoined it to circumvent all but about 40 centimetres. Known as a jejunioileal bypass, it caused remarkable weight loss but also an array of unpleasant side effects, including bloating, diarrhoea, anal burning and dehydration. Bacterial populations in the bypassed intestine continually rose and the liver became inflamed. "Everybody realized that five years after you have this, you lose your liver," says David Cummings, an endocrinologist at the University of Washington in Seattle.

Today's gold standard is the Roux-en-Y gastric bypass. Pioneered in 1977, the procedure creates a small pouch at the top of the stomach and reroutes the small intestine to connect to it. The bypassed section gets reconnected to the intestine, forming a 'Y' shape, so that it can still drain fluids and bacteria, reducing the risk of festering growth.

Even in the early days of gastric bypass, surgeons noticed that the operation had swift effects on metabolism: patients' blood-sugar levels normalized within a week or so. "We were surprised by the rapidity of the improvement," read a 1987 study reporting on 397 procedures³, "even though the patients were still clearly morbidly obese."

Patients said that they were not as hungry as before the surgery, and that they ate fewer meals and snacked less. Over time, their food preferences seemed to change, too; anecdotal reports suggested that they often chose salads over desserts and fatty foods. These shifts could not be explained by reduced stomach size alone, Cummings notes — if the reason

ILLUSTRATION BY MARCUS BUTT



was mechanical, patients would simply eat lots of small meals. “That got the field wondering, what’s going on with hunger, here?”

In 2002, Cummings and his colleagues identified one of the first biochemical markers associated with the bypass. They had tracked blood levels of ghrelin, the ‘hunger hormone’ produced by cells in the gastrointestinal tract, in more than two dozen people. Normally, ghrelin levels rise sharply when the stomach is empty and then drop after a meal. Surgery suppressed these fluctuations, Cummings found⁴. The normal peaks and valleys of ghrelin production went pancake flat. “It’s pretty dramatic,” he says.

But getting a better handle on the mechanisms required an animal model. Lee Kaplan, director of the Massachusetts General Hospital Weight Center in Boston, looked to rats — a daunting task given their tiny innards. He recruited a young surgeon from Greece, Nicholas Stylopoulos, and the duo, along with a few other research groups, began to publish papers on what happened to the animals after surgery. The research has shown that just like in people, bypass surgery stabilizes glucose levels⁵, boosts metabolism⁶ and steers the animals to choose low-fat over high-fat meals⁷.

GUT MICROBES

A potential explanation could lie in the trillions of microbes that reside in the gut. In 2009, Rosa Krajmalnik-Brown from Arizona State University in Tempe and her colleagues sequenced the bacterial genes present in

faeces from three people who had received a gastric bypass. Compared with obese and normal-weight controls, their guts contained proportionally fewer bacteria from the usually abundant Firmicutes phylum, and excess levels of the Gammaproteobacteria class⁸. “Even with that small sample size we were able to get statistically significant differences because the microbiota changed so drastically,” Krajmalnik-Brown says.

The researchers do not know why these particular changes occurred, but they say it could be because Firmicutes die when oxygen is present, and shortening the gastrointestinal tract means that oxygen that is normally consumed in the small intestine reaches the colon. Alternatively, the changes could occur because food is being digested faster. (The group did not test microbial make-up in individuals before surgery, but is now working on a follow-up study that compares before and after.) A similar shift in gut flora has been reported in rats undergoing a gastric bypass⁹.

Whether this bacterial shift drives a change in health is hard to say, but there are some indications that the microbes contribute to metabolic changes. Kaplan and his colleagues performed a gastric bypass on obese mice, then transplant the altered gut bacteria into mice bred to be microbe-free. These recipient mice were not obese, but still lost about 5% of their weight after the transplant¹⁰ (see *Nature* <http://doi.org/tjq; 2013>).

This research and other strands of

evidence suggest that metabolic regulation could begin in the gut, which has the ability to send messages to the brain, liver, pancreas, kidneys and immune system. “The idea that a lot of the information starts at the gut is a relatively new concept,” says Kaplan.

For example, researchers have now found that bile acids have a role in signalling. These fluids help to emulsify fats so that the lipids are metabolized more efficiently, but they also act as hormones, signalling to receptors in the gut. Randy Seeley, a neuroscientist at the University of Michigan Health System in Ann Arbor, and his colleagues decided to look at what happens when one of these bile-activated-receptors — the farnesoid-X receptor (FXR), which helps to regulate glucose metabolism — is deleted in mice.

The researchers overfed both mutant and control mice until they were fat, and then did a vertical sleeve gastrectomy. (This procedure shrinks the stomach like a gastric bypass does, but does not circumvent any of the small intestine.) A week after surgery, both types of mice lost a lot of weight. By the fifth week, however, only the control mice had managed to keep it off; the mutants had gained it all back¹¹. Without FXR and the messages carried by bile acids, the surgery fails to work.

Intriguingly, the control mice, but not the mutants, showed a notable increase in the abundance of *Roseburia*, a Firmicutes bacterium that tends to be suppressed in people

with diabetes, suggesting that FXR and its related biological pathways could turn out to be therapeutic targets in this disease.

Bile-acid and bacterial changes could affect the gut's communication with organs responsible for the glucose dysregulation that causes diabetes. But a study published last year¹² suggests that the gut itself shows changes in glucose metabolism after surgery (see *Nature* <http://doi.org/tjr; 2013>).

Using a rat model of gastric bypass, Stylopoulos, who now runs his own laboratory at Boston Children's Hospital, and his colleagues showed that the 'Roux limb' — the piece of intestine that runs from the stomach pouch to the reconnected intestine — expands dramatically in width and length after surgery. "It really doubles in size," Stylopoulos says, and it stays that way. That makes sense, because without a full-sized stomach, the tissue must adapt to heaps of undigested food. But the limb's rapid growth requires a lot of energy, which comes from glucose. The changing organ starts to use more glucose, and the change is maintained over time, Stylopoulos says. "Essentially, the intestine becomes a bigger and a more hungry organ that needs more glucose than before."

Stylopoulos believes that this tissue growth in the gut is the main driver of the surgery's remarkable metabolic benefits — not the reduction in calorie intake. "Surgery works because it changes the physiology," he says.

Weight loss is still important, however, because it triggers a series of changes that help to curb diabetes.

PROBLEMS IN TRANSLATION

How well do these findings translate into people? "These are elegant studies," says Samuel Klein, director of the Center for Human Nutrition at Washington University School of Medicine in St Louis, Missouri. But, he asks: "Is the bariatric surgical procedure in a rodent the same as in a human?"

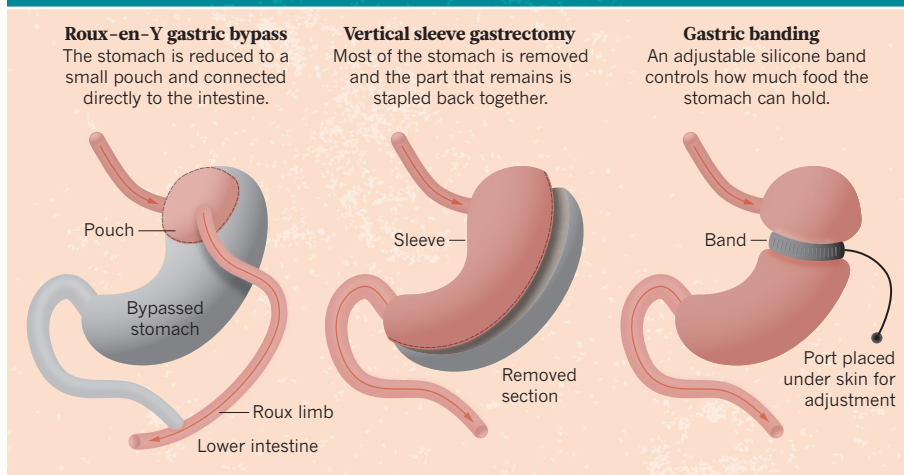
Klein allows that just like rodents, people have a marked improvement in blood-glucose regulation within days of bypass surgery. But that could be because their caloric intake goes from around 4,000 calories a day to just 400, he says. "Anyone who has abdominal surgery is not going to be very hungry after the operation."

Rates of diabetes remission are much higher after gastric bypass than after gastric banding — in which a silicone band squeezes around the stomach to restrict the flow of food (see 'Surgical selection'). Animal studies suggest that that is because the bypass alters metabolism in a way that banding does not, but Klein believes that it is simply because people who have a bypass tend to lose much more weight.

To probe this, Klein compared people who had lost one-fifth of their weight after gastric bypass with those who lost the same

Surgical selection

A candidate for bariatric surgery can typically choose from three broad categories. All procedures reduce the amount of food the person can eat, but bypass and gastrectomy have the strongest effects on weight loss and other ailments such as diabetes.



amount with banding. All patients showed dramatic improvements in glucose tolerance, insulin sensitivity and the function of pancreatic β -cells, which release insulin¹³. "We did not see any hint" of differences between the groups, he says. The major caveat of this study is that none of the volunteers had diabetes, so Klein's group is now repeating the study in people with the disease. "It could be a whole different ball game," he says.

Still, he agrees that rodent studies provide a relatively quick way to investigate specific biological pathways and test hypotheses about why only some procedures curb diabetes and why certain patients are more likely to benefit than others. By testing individual pathways, researchers hope that they can develop personalized treatments — whether drugs, probiotics or lifestyle changes — that change the specific pathway that has gone awry in a patient.

For Courcoulas, the variability and unpredictability in patient response — in both weight loss and diabetes remission — is the most important issue that animal studies could address. When talking to prospective patients about their surgical options, she frequently refers to a study she published last year¹⁴ that tracked nearly 2,500 people who had undergone various types of bariatric surgery.

After three years, those who received gastric banding had lost, on average, about 16% of their weight, whereas those who had a gastric bypass lost 32%. Banding also led to partial remission of diabetes in 29% of people, compared with 68% for bypass. In general, Courcoulas notes, most people lose a lot of weight in the first six months, irrespective of the procedure. But after that, they diverge wildly: some people continue to lose at a rapid clip, others plateau and still others gain some back. This uncertainty partly

explains why so few people who are eligible for surgery choose to have it, she says. At her centre, nearly 1,500 people a year attend group informational sessions to learn the basics of weight-loss surgery. Only 1,000 of them will elect to talk to a surgeon, and 700 will go on to have an operation.

"The big question is, what are the factors, the predictors for someone's success after surgery?" Courcoulas says. Clinical studies have identified some contributors — iron deficiency, liver fibrosis and being older than 50 years, for instance, are all associated with less weight loss¹⁵. But none of these is absolute. The only thing clear, Courcoulas says, is the need to identify better biological markers. "My colleagues in basic science," she says, "are going to be making a big contribution in doing that." ■

Virginia Hughes is a science journalist based in New York.

- Adams, T. D. *et al.* *J. Am. Med. Assoc.* **308**, 1122–1131 (2012).
- Henrikson, V. *Nordisk Medicin* **47**, 744 (1952). (republished in *Obes. Surg.* **4**, 54–55; 1994.)
- Pories, W. J., Caro, J. F., Flickinger, E. G., Meelheim, H. D. & Swanson, M. S. *Ann. Surg.* **206**, 316–323 (1987).
- Cummings, D. E. *et al.* *N. Engl. J. Med.* **346**, 1623–1630 (2002).
- Rubino, F. & Marescaux, J. *Ann. Surg.* **239**, 1–11 (2004).
- Stylopoulos, N., Hoppin, A. G. & Kaplan, L. M. *Obesity* **17**, 1839–1847 (2009).
- Zheng, H. *et al.* *Am. J. Physiol.* **297**, R1273–R1282 (2009).
- Zhang, H. *et al.* *Proc. Natl. Acad. Sci. USA* **106**, 2365–2370 (2009).
- Li, J. V. *et al.* *Gut* **60**, 1214–1223 (2011).
- Liou, A. P. *et al.* *Sci. Transl. Med.* **5**, 178ra41 (2013).
- Ryan, K. K. *et al.* *Nature* **509**, 183–188 (2014).
- Saeidi, N. *et al.* *Science* **341**, 406–410 (2013).
- Bradley, D. *et al.* *J. Clin. Invest.* **122**, 4667–4674 (2012).
- Courcoulas, A. P. *et al.* *J. Am. Med. Assoc.* **310**, 2416–2425 (2013).
- Still, C. D. *et al.* *Obesity* **22**, 888–894 (2014).

COMMENT

MENTAL HEALTH Collaborative research needed to improve psychological treatments **p.287**



VACCINES The dramatic tale of two typhus biologists under the Nazis **p.291**

CAREERS Virtual mobility will drive equal opportunities for scientists in Europe **p.292**

EDUCATION Top graduates volunteer to teach China's rural poor **p.292**

ASIANET-PAKISTAN/ALAMY



A health worker gives a dose of polio vaccine to a child in Chaman, Pakistan, near the Afghan border, in May.

Polio eradication hinges on child health in Pakistan

Boosting basic medical services and routine immunizations — not travel vaccinations — is the key to ending polio worldwide, says **Zulfiqar Ahmed Bhutta**.

Until about a year ago, a world free of poliomyelitis seemed to be imminent. In 1988, about 350,000 people in 125 countries became paralysed by the virus. Last year, only 406 cases were reported, with 160 of them in just a few areas of the three countries where polio remains endemic: Afghanistan, Nigeria and Pakistan. In April 2013, charities and governments pledged US\$4 billion to a six-year plan developed by the World Health Organization (WHO) to eradicate polio. In March, after India had gone three years with no new

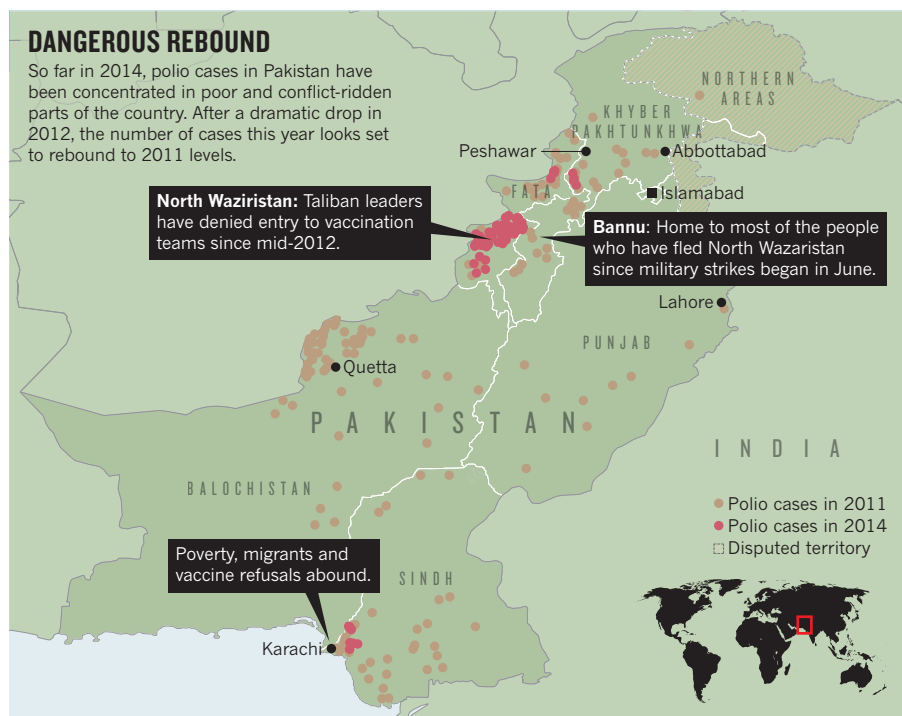
cases, the WHO certified its southeast Asia region (which does not include Afghanistan and Pakistan) as polio-free.

But in May, the WHO declared polio an international public-health emergency, particularly because of the high risk of international spread from Pakistan, Cameroon and Syria (see go.nature.com/7z3efj). Disrupted vaccination programmes in war-torn places are partly to blame.

Confronted by this, the WHO took an unprecedented step: it called for mandatory polio vaccination for everyone travelling to

or from Pakistan, Syria and Cameroon, and encouraged travel vaccinations for Afghanistan, Nigeria and others¹. Formal international travel restrictions for Pakistan began on 1 June. Analyses in the past few years show² that symptom-free adults transmit polio at surprisingly high rates. However, computer modelling described³ earlier this month suggests that immunizing adults to control an outbreak is less effective than previously believed.

In my view, vaccinating travellers will be ineffective and it could make polio ►



► harder to eliminate in the poor and conflict-ridden parts of Pakistan. It is largely here that the final battle to eradicate polio from the world will be won or lost.

Cases of polio in Pakistan increased from 18 in the first six months of 2013 to 88 in the first half of 2014 (ref. 4). Of these, 75% were in the regions known as the Federally Administered Tribal Areas (FATA) in the northwest (see 'Dangerous rebound'). Here, access for polio-vaccination teams is severely restricted by conflict and insecurity.

Since mid-June, the situation has worsened. In the wake of government military action against Taliban insurgents, more than 800,000 people from Waziristan in the FATA have been displaced to neighbouring parts of Pakistan and Afghanistan. Instead of focusing on the vaccination of international travellers, Pakistan, the WHO and immunization services should provide immediate health care to displaced families and others in these high-risk areas.

PRECIOUS DOSES

Federal and provincial governments in Pakistan have scrambled to set up vaccination points at all ports and airports, and at more than 130 public hospitals. The government of Punjab, Pakistan's richest and most populous province, also rushed to impose vaccination requirements for the main routes of entry. The federal government made polio vaccination mandatory at major entry and exit points in the FATA, especially in North Waziristan, although much of the long, troubled border with Afghanistan is unpatrolled.

Official sources estimate that more than 10 million doses are needed just for the

air travellers entering or exiting Pakistan each year, including the roughly 7 million Pakistani citizens who work overseas, mostly as labourers in the Middle East. The donor community has provided 200,000 doses of injectable polio vaccine for refugees, but no further financial support has been pledged for more doses or for trained staff to perform vaccinations and issue certificates to adult travellers at public hospitals.

So far, the only service offered for free to travellers is the oral vaccine from the supplies of national polio programmes. (Some 300 million doses of oral polio vaccine, mostly furnished by the United Nations children's charity UNICEF, are needed annually to vaccinate young children in Pakistan.) Pakistan's army requested 60,000 doses of inactivated injectable polio vaccine as a priority for its troops. Adults must buy this type of vaccine privately at a cost of \$4.30 per dose — a huge expense in an area where the average monthly income is about \$100. Newspapers report that getting a vaccination certificate is as difficult and expensive as getting a visa. An industry of fake certification could emerge.

There is no precedent to predict how well these travel restrictions will work. I travelled out of Karachi airport on 6 and 15 June. Although vaccination counters had been set up, I saw no queues of travellers waiting to receive polio vaccines, and no one asked me for a vaccination card at any of the multiple checkpoints. Furthermore, polio transmission from Pakistan to Afghanistan occurs mostly across an unregulated border.

Meanwhile, Pakistan's efforts to vaccinate young children have fallen behind. Some of the blame can be pinned on the ill-planned

abolition of its ministry of health in 2011 and the subsequent devolution of health services to the provinces. Although the ministry was reinstated last year and federal polio efforts are now back in operation, they are still weak.

That said, Pakistan deserves much more credit than it has received for its past work to eradicate polio, especially in its troubled tribal regions: it has staged more than 130 national and regional polio-immunization efforts since it began house-to-house vaccination campaigns in 2000.

But the emphasis on polio, to the neglect of other health services, has long fuelled beliefs that polio immunization is an external initiative operating for outsiders' benefit. Anti-Western sentiment has led to repeated attacks on polio-eradication workers, volunteers and security personnel; more than 80 have been killed since December 2012. This year, polio teams have been hit by roadside bombs and by gunmen on motorcycles. In March, a Pakistani polio worker was kidnapped and shot.

Resistance to polio campaigns is more entrenched and violent in Pakistan than in most other countries. Disastrously, mobile-vaccination teams came under more suspicion than ever⁵ after it emerged that the US Central Intelligence Agency had staged a fake hepatitis B vaccination project in the Pakistani city of Abbottabad in 2011 to try to trace Osama bin Laden.

Although international Islamic scholars have spoken up for polio eradication, support for it from local religious and society leaders on the ground has been, at best, lukewarm. In the 1980s and 90s, warring factions in Latin America and in Africa agreed to 'days of tranquility' to permit mass polio immunizations. In Pakistan, by contrast, a handful of Taliban leaders in the tribal areas of North Waziristan and the Khyber Agency have, since mid-2012, denied entry to vaccination teams as a protest against US drone strikes.

In May this year, the Pakistani army moved to provide security to vaccination teams in the FATA, but it has not offered support to other mainstream health workers. This and the hastily imposed travel regulations will only give credence to claims that polio eradication is part of a foreign agenda.

PRESCRIPTION PACKAGE

Providing polio vaccines as part of a package of health services is a better way to engage local communities and religious leaders than through a narrow, polio-specific programme. Nigeria and Afghanistan have made remarkable progress in reaching difficult populations in this way, and cases dropped by about 60% in both nations from 2012 to 2013⁴. The Taliban do not actively keep children from being immunized for measles or from receiving care for diarrhoea or malnutrition.

Currently, Pakistan has one of the highest rates of child mortality in south Asia⁶.

Children face much bigger health threats than polio. But immunization services for major childhood diseases such as diphtheria, tetanus and measles remain plagued with inefficiencies, poor oversight and a shortage of resources.

Full immunization rates for children in the country were last year estimated at 54% with wide variations across the country⁷, compared to more than 95% in nearby Bangladesh. The figures for Pakistan may even be an overestimate: the survey excluded the FATA and vulnerable populations in mega-cities. In a household survey conducted this year, my colleagues and I found that 25% of children under five years in the urban slums of Karachi were not vaccinated for any childhood disease; the same was true for 64% of children in a relatively peaceful district of the FATA.

The time to act is now. The military offensive in North Waziristan has, paradoxically, opened up opportunities to provide health services to children from the FATA through care for displaced families. This could contribute to building community support and to re-establish the rule of law in conflict-ridden areas once people return. Ongoing support will be necessary to eradicate polio: children require multiple doses of vaccine to build immunity.

I fervently hope that the government and concerned agencies will devote their energies to scaling up full immunization efforts in these displaced and marginal populations, rather than diverting resources to international travellers. This is a chance to eradicate polio from the planet. ■

Zulfiqar Ahmed Bhutta is director of the Center of Excellence in Women and Child Health at the Aga Khan University in Karachi, Pakistan, and co-director of the Sick Kids Center for Global Child Health in Toronto, Canada.
e-mail: zulfiqar.bhutta@aku.edu

1. Rutter, P. D. & Donaldson, L. J. *Lancet* **383**, 1695–1697 (2014).
2. Mayer, B. T. et al. *Am. J. Epidemiol.* **177**, 1236–1245 (2013).
3. Blake, I. M. et al. *Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1323688111> (2014).
4. Independent Monitoring Board of the Global Polio Eradication Initiative *Ninth Report* — May 2014 (Global Polio Eradication Initiative, 2014).
5. Bhutta, Z. A. *Br. Med. J.* **346**, f280 (2013).
6. Bhutta, Z. A. et al. *Lancet* **381**, 2207–2218 (2013).
7. National Institute of Population Studies and ICF International *Pakistan Demographic and Health Survey 2012–13* (NIPS & ICF International, 2013).

The views expressed are those of the author alone, and do not necessarily reflect those of any institution with which he is affiliated. The author declares competing financial interests: see go.nature.com/cnhjcu for details.



A call for mental-health science

Clinicians and neuroscientists must work together to understand and improve psychological treatments, urge **Emily A. Holmes, Michelle G. Craske and Ann M. Graybiel.**

How does one human talking to another, as occurs in psychological therapy, bring about changes in brain activity and cure or ease mental disorders? We don't really know. We need to.

Mental-health conditions, such as post-traumatic stress disorder (PTSD),

obsessive-compulsive disorder (OCD), eating disorders, schizophrenia and depression, affect one in four people worldwide. Depression is the third leading contributor to the global burden of disease, according to the World Health Organization. Psychological treatments have been subjected ▶

► to hundreds of randomized clinical trials and hold the strongest evidence base for addressing many such conditions. These activities, techniques or strategies target behavioural, cognitive, social, emotional or environmental factors to improve mental or physical health or related functioning. Despite the time and effort involved, they are the treatment of choice for most people (see ‘Treating trauma with talk therapy’).

For example, eating disorders were previously considered intractable within our life time. They can now be addressed with a specific form of cognitive behavioural therapy (CBT)¹ that targets attitudes to body shape and disturbances in eating habits. For depression, CBT can be as effective as antidepressant medication and provide benefits that are longer lasting². There is also evidence that interpersonal psychotherapy (IPT) is effective for treating depression.

A HOUSE DIVIDED

But evidence-based psychological treatments need improvement. Although the majority of patients benefit, only about half experience a clinically meaningful reduction in symptoms or full remission, at least for the most common conditions. For example, although response rates vary across studies, about 60% of individuals show significant improvement after CBT for OCD, but nearly 30% of those who begin therapy do not complete it³. And on average, more than 10% of those who have improved later relapse⁴. For some conditions, such as bipolar disorder, psychological treatments are not effective or are in their infancy.

Moreover, despite progress, we do not

yet fully understand how psychological therapies work — or when they don’t. Neuroscience is shedding light on how to modulate emotion and memory, habit and fear learning. But psychological understanding and treatments have, as yet, profited much too little from such developments.

It is time to use science to advance the psychological, not just the pharmaceutical, treatment of those with mental-health problems. Great strides can and must be made by focusing on concerns that are common to fields from psychology, psychiatry and pharmacology to genetics and molecular biology, neuroscience, cognitive and social sciences, computer science, and mathematics. Molecular and theoretical scientists need to engage with the challenges that face the clinical scientists who develop and deliver psychological treatments, and who evaluate their outcomes. And clinicians need to get involved in experimental science. Patients, mental-health-care providers and researchers of all stripes stand to benefit.

Interdisciplinary communication is a problem. Neuroscientists and clinical scientists meet infrequently, rarely work together, read different journals, and know relatively little of each other’s needs and discoveries. This culture gap in the field of mental health has widened as brain science has exploded. Researchers in different disciplines no longer work in the same building, let alone the same department, eroding communication. Separate career paths in neuroscience, clinical psychology and psychiatry put the fields in competition for scarce funding.

Part of the problem is that for many people, psychological treatments still conjure up

notions of couches and quasi-mystical experiences. That evidence-based psychological treatments target processes of learning, emotion regulation and habit formation is not clear to some neuroscientists and cell biologists. In our experience, many even challenge the idea of clinical psychology as a science and many are unaware of its evidence base. Equally, laboratory science can seem abstract and remote to clinicians working with patients with extreme emotional distress and behavioural dysfunction.

CHANGING ATTITUDES

Research on psychological treatments is, in the words of this journal, “scandalously under-supported” (see *Nature* **489**, 473–474; 2012). Mental-health disorders account for more than 15% of the disease burden in developed countries, more than all forms of cancer. Yet it has been estimated that the proportion of research funds spent on mental health is as low as 7% in North America and 2% in the European Union.

Within those slender mental-health budgets, psychological treatments receive a small slice — in the United Kingdom less than 15% of the government and charity funding for mental-health research, and in the United States the share of National Institute of Mental Health funding is estimated to be similar. Further research on psychological treatments has no funding stream analogous to investment in the pharmaceutical industry.

This Cinderella status contributes to the fact that evidence-based psychological treatments, such as CBT, IPT, behaviour therapy and family therapy, have not yet fully benefitted from the range of dramatic advances in the neuroscience related to emotion, behaviour and cognition. Meanwhile, much of neuroscience is unaware of the potential of psychological treatments. Fixing this will require at least three steps.

THREE STEPS

Uncover the mechanisms of existing psychological treatments. There is a very effective behavioural technique, for example, for phobias and anxiety disorders called exposure therapy. This protocol originated in the 1960s from the science of fear-extinction learning and involves designed experiences with feared stimuli. So an individual who fears that doorknobs are contaminated might be guided to handle doorknobs without performing their

“Neuroscientists and clinical scientists meet infrequently.”

CASE STUDY

Treating trauma with talk therapy

Ian was filling his car with petrol and was caught in the cross-fire of an armed robbery. His daughter was severely injured. For the following decade Ian suffered nightmares, intrusive memories, flashbacks of the trauma and was reluctant to drive — symptoms of post-traumatic stress disorder (PTSD).

Ian had twelve 90-minute sessions of trauma-focused cognitive behavioural therapy, the treatment with the strongest evidence-base for PTSD, which brings about improvement in about 75% of cases. As part of his therapy, Ian was asked to replay the traumatic memory vividly in his mind’s eye. Ian also learned that by

avoiding reminders of the trauma his memories remained easily triggered, creating a vicious cycle. Treatment focused on breaking this cycle by bringing back to his mind perceptual, emotional and cognitive details of the trauma memory.

After three months of treatment, Ian could remember the event without being overwhelmed with fear and guilt. The memory no longer flashed back involuntarily and his nightmares stopped. He began to drive again.

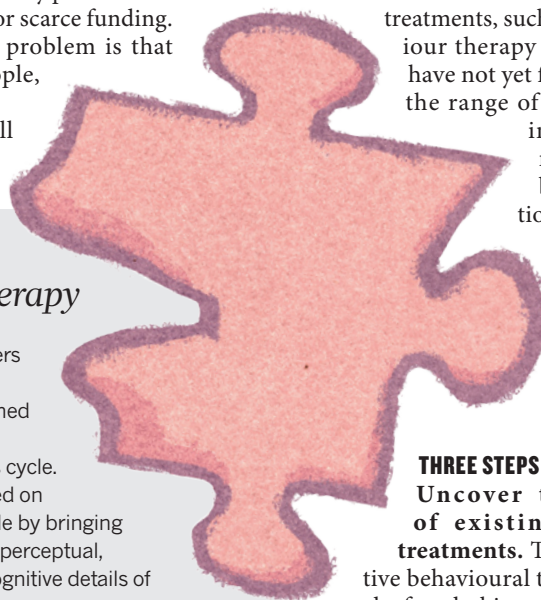


ILLUSTRATION BY DAVID PARKINS

compulsive cleansing rituals. They learn that the feared stimulus (the doorknob) is not as harmful as anticipated; their fears are extinguished by the repeated presence of the conditional stimulus (the doorknobs) without safety behaviours (washing the doorknobs, for example) and without the unconditional stimulus (fatal illness, for example) that was previously signalled by touching the doorknob.

But in OCD, for instance, nearly half of the people who undergo exposure therapy do not benefit, and a significant minority relapse. One reason could be that extinction learning is fragile — vulnerable to factors such as failure to consolidate or generalize to new contexts. Increasingly, fear extinction is viewed⁵ as involving inhibitory pathways from a part of the brain called the ventromedial prefrontal cortex to the amygdala, regions of the brain involved in decision-making, suggesting molecular targets for extinction learning. For example, a team led by one of us (M.G.C.), a biobehavioural clinical scientist at the University of California, Los Angeles, is investigating the drug scopolamine (usually used for motion sickness and Parkinson's disease) to augment the generalization of extinction learning in exposure therapy across contexts. Others are trialling D-cycloserine (originally used as an antibiotic to treat tuberculosis) to enhance the response to exposure therapy⁶.

Another example illustrates the power of interdisciplinary research to explore cognitive mechanisms. CBT asserts that many clinical symptoms are produced and maintained by dysfunctional biases in how emotional information is selectively attended to, interpreted and then represented in memory. People who become so fearful and anxious about speaking to other people that they avoid eye contact and are unable to attend their children's school play or a job interview might notice only those people who seem to be looking at them strangely (negative attention bias), fuelling their anxiety about contact with others. A CBT therapist might ask a patient to practice attending to positive and benign faces, rather than negative ones.

In the past 15 years, researchers have discovered that computerized training can also modify cognitive biases⁷. For example, asking a patient (or a control participant) to repeatedly select the one smiling face from a crowd of frowning faces can induce a more positive attention bias. This approach enables researchers to do several things: test

the degree to which a given cognitive bias produces clinical symptoms; focus on how treatments change biases; and explore ways to boost therapeutic effects.

One of us (E.A.H.) has shown with colleagues that computerized cognitive bias modification alters activity in the lateral prefrontal cortex⁸, part of the brain system that controls attention. Stimulating neural activity in this region electrically augments the computer training. Such game-type tools offer the possibility of scalable, 'therapist-free' therapy.

Optimize psychological treatments and generate new ones. Neuroscience is providing unprecedented information about processes that can result in, or relieve, dysfunctional behaviour. Such work is probing the flexibility of memory storage, the degree to which emotions and memories can be dissociated, and the selective neural pathways that seem to be crucial for highly specialized aspects of the emotional landscape and can be switched on and off experimentally. These advances can be translated to the clinical sphere.

For example, neuroscientists (including A.M.G.) have now used optogenetics to block⁹ and produce¹⁰ compulsive behaviour such as excessive grooming by targeting different parts of the orbitofrontal cortex. The work was inspired by clinical observations that OCD symptoms, in part, reflect an over-reaction to conditioned stimuli in the environment (the doorknobs in the earlier example). These experiments suggest that a compulsion, such as excessive grooming, can be made or broken in seconds through targeted manipulation of brain activity. Such experiments, and related work turning on and off 'normal' habits with light that manipulates individual cells (optogenetics), raise the tantalizing possibility of optimizing behavioural techniques to activate the brain circuitry in question.

Forge links between clinical and laboratory researchers. We propose an umbrella discipline of mental-health science that joins behavioural and neuroscience approaches to problems including improving psychological treatments. Many efforts are already being made, but we need to galvanize the next generation of clinical scientists and neuroscientists to interact by creating career opportunities that enable them to experience advanced methods in both.

New funding from charities, the US National Institutes of Health and the European framework Horizon 2020 should strive to maximize links between fields. A positive step was the announcement in February by the US National Institute of Mental Health that it will fund only the psychotherapy

trials that seek to identify mechanisms.

Neuroscientists and clinical scientists could benefit enormously from national and international meetings. The psychological treatments conference convened by the mental-health charity MQ in London in December 2013 showed us that bringing these groups together can catalyse new ideas and opportunities for collaboration. (The editor-in-chief of this journal, Philip Campbell, is on the board of MQ.) Journals should welcome interdisciplinary efforts — their publication will make it easier for hiring committees, funders and philanthropists to appreciate the importance of such work.

WHAT NEXT

By the end of 2015, representatives of the leading clinical and neuroscience bodies should meet to hammer out the ten most pressing research questions for psychological treatments. This list should be disseminated to granting agencies, scientists, clinicians and the public internationally.

Mental-health charities can help by urging national funding bodies to reconsider the proportion of their investments in mental health relative to other diseases. The amount spent on research into psychological treatments needs to be commensurate with their impact. There is enormous promise here. Psychological treatments are a lifeline to so many — and could be to so many more. ■

Emily A. Holmes is at the Medical Research Council Cognition & Brain Sciences Unit, Cambridge, UK, and in the Department for Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden. **Michelle G. Craske** is in the Department of Psychology and Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, USA. **Ann M. Graybiel** is in the Department of Brain and Cognitive Sciences, McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. e-mail: emily.holmes@mrc-cbu.cam.ac.uk

1. Fairburn, C. G. *et al.* *Am. J. Psychiatry* **166**, 311–319 (2009).
2. Hollon, S. D. *et al.* *Arch. Gen. Psychiatry* **62**, 417–422 (2005).
3. Foa, E. B. *et al.* *Am. J. Psychiatry* **162**, 151–161 (2005).
4. Simpson, H. B. *et al.* *Depress. Anxiety* **19**, 225–233 (2004).
5. Vervliet, B., Craske, M. G. & Hermans, D. *Annu. Rev. Clin. Psychol.* **9**, 215–248 (2013).
6. Otto, M. W. *et al.* *Biol. Psychiatry* **67**, 365–370 (2010).
7. MacLeod, C. & Mathews, A. *Annu. Rev. Clin. Psychol.* **8**, 189–217 (2012).
8. Browning, M., Holmes, E. A., Murphy, S. E., Goodwin, G. M. & Harmer, C. J. *Biol. Psychiatry* **67**, 919–925 (2010).
9. Burguière, E., Monteiro, P., Feng, G. & Graybiel, A. M. *Science* **340**, 1243–1246 (2013).
10. Ahmari, S. E. *et al.* *Science* **340**, 1234–1239 (2013).

HISTORY OF MEDICINE

Typhus and tyranny

Tilli Tansey ponders a turbulent history of vaccine research in Nazi-occupied Europe.

Lice thrive in war. Overcrowded conditions, the large-scale movements of troops and displaced persons, and the breakdown of rudimentary hygiene are ideal for the survival and transmission of body lice (*Pediculus humanus humanus*) and their sinister bacterial loads: *Rickettsia prowazekii*, the cause of the deadly disease typhus. In 1918, more than 650,000 cases of typhus were recorded in newly independent Poland alone.

As writer Arthur Allen relates in *The Fantastic Laboratory of Dr. Weigl*, it was this potent mix of geopolitical chaos and rampant disease that sealed the fates of two Polish biologists: Rudolf Weigl and Ludwik Fleck.

Weigl's laboratory in Lwów, Poland — now Lviv in Ukraine — is little remembered. But between the first and second world wars, it was a world centre of typhus-vaccine research. With government support, Weigl was the first to culture *Rickettsia* by harnessing lice as experimental animals. He devised an anal-inoculation technique to infect the insects with the bacteria, and marshalled human volunteers to nourish them. The typhus-engorged midguts of the lice were the raw material for vaccine production, and by the early 1930s the first reliable typhus vaccine was being tested and distributed.

In neighbouring Germany, Nazi propaganda associated lice with Jews, so in peace time there had been little interest in producing a vaccine. Priorities changed as the Second World War progressed and German troops first invaded, and were later defeated in, typhus-ridden territories in central and Eastern Europe.

In German-occupied Poland in 1941, Weigl's lab was put under the control of the Nazi armed forces. He and — elsewhere in Lwów — his former assistant, Fleck, were ordered to develop and produce typhus vaccines.

Weigl's lab became the town's intellectual centre, replacing pre-war cafe culture. Dismissed academics, many of them Jewish, applied to become louse feeders, earning a tiny monthly payment and wider protection from looters and attackers as stories circulated of clothes and homes crawling with lice. As the insects, in boxes strapped to the feeders' legs, sucked up blood, their nurturers would discourse on topics as diverse as mathematics,



Polish microbiologist Rudolf Weigl in his typhus laboratory during the Second World War.

philosophy and psychology. Feeders were trained not to scratch their irritated skin, to prevent infection — not of themselves but of the valuable lice.

Unsurprisingly in an institute in occupied territory, charged with producing life-saving medicines for the enemy, tensions arose. Weigl's scientific pride in producing a perfect vaccine contrasted with the desire among much of his staff to disrupt production.

It was a taut equation: if supply to the German army was seriously affected, the institute would draw unwelcome attention from the Gestapo, who might close the lab, or take it over. The vaccine workers continued as best they could. Occasional subterfuges created suboptimal vaccines, and a deal allowed a small amount of vaccine for

private use, which, it is claimed, found its way to Warsaw's Jewish ghetto.

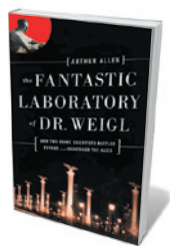
Fleck was in a very different situation. As a Jew, he was consigned to the Lwów ghetto after German occupation. He was arrested in February 1943, and thereafter worked in labs in concentration and extermination camps, under the direct control of the SS. At Auschwitz and especially at Buchenwald, he and colleagues devised another solution to the problem of working for the enemy. Fleck cultured *Rickettsia* in experimental animals, mainly rabbits, and then harvested the animals' lungs. From these his team manufactured a useless vaccine. Untrained co-workers and ignorant SS supervisors unknowingly supported the pretence.

After the war, the new world order in Poland treated both men harshly. Despite being appointed to a university chair in Krakow, and offered vaccine-making facilities in Moscow, Weigl found his Nazi associations and refusal to become involved with the socialist regime damaging. He died, broken and forgotten, in 1957. Fleck worked in Lublin and then Warsaw, increasingly subject to anti-Semitism. In the year of Weigl's death, he emigrated to Israel, where he worked in bacteriology until his own death in 1961.

Perhaps Fleck, now known for his 1935 *The Genesis and Development of a Scientific Fact*, was bold in attempting his large-scale deception. But was Weigl, when he extended a warm welcome to his new German masters, being a disinterested scientist, a subtle subversive or a genuine sympathizer? Allen argues for heroic rebellion, citing Israel's recognition of Weigl with an honour awarded to those who risked their lives to save Jews during the Holocaust. That view has been challenged by historians such as Paul Weindling, and even Allen's own account is at times ambiguous.

The book's style and purpose are perplexing. It has some of the trappings of an academic work, such as multi-language references, but words such as "claptrap" jar with the predominantly scholarly tone of the text. All in all, however, fascinating stories emerge. ■

Tilli Tansey is professor of the history of modern medical sciences at Queen Mary, University of London.
e-mail: t.tansey@qmul.ac.uk



The Fantastic Laboratory of Dr. Weigl: How Two Brave Scientists Battled Typhus and Sabotaged the Nazis
ARTHUR ALLEN
W. W. Norton: 2014.

EAST NEWS POLAND

Correspondence

Editorials should heed social scientists

Your Editorials consistently recommend that policy decisions should be backed by sound science, including social science. However, my subjective analysis of a sample of these articles indicates that you do not always follow your own advice.

Roughly half of the 141 Editorials you published in the year from October 2012 relate to policy issues. Of these, only about 10% use literature citations to support their arguments. By contrast, 35% of the Editorials that express ideas with no direct bearing on policy are backed up by referencing.

Moreover, your policy proposals sometimes contradict the consensus opinion among social scientists. Take climate-change mitigation: you tend to highlight piecemeal emissions-reduction policies, such as the introduction of fuel standards, building codes or subsidies for renewable energy. However, most economists dismiss government micromanagement of polluting activities as inefficient and unfair, and would prefer to see the establishment of a universal carbon tax (see N. G. Mankiw *East. Econ. J.* **35**, 12–23; 2009; and go.nature.com/ylxraf).

You also seem to overlook the diversity of opinion among social scientists on topical issues. For instance, you frequently make a plea for more power for the US Food and Drug Administration, without acknowledging the debate over whether the social good would be better served by increasing or decreasing the agency's regulatory power (see go.nature.com/pxxswg).

And in your persistent request for more government money for research, you could make a stronger case by using tools devised by social scientists to estimate the optimal size and allocation of science budgets.

Nature's views probably coincide with the default views of its readership and the public.

This should not distract you from publicizing instances of “sound science and evidence on a matter of public interest” (*Nature* **491**, 160; 2012).

Marcelino Fuentes *University of A Coruña, Spain.*
marcelinofuentes@gmail.com

Virtual mobility can drive equality

At a EuroScience Open Forum meeting last month, scientists, policy-makers and the public discussed ‘virtual mobility’. Could it replace the conventional geographical mobility of early-career researchers between labs? (See also R. Garwood *Nature* **510**, 313; 2014.)

The group concluded that virtual mobility would work, but should be combined with short-term visits to other labs to allow face-to-face contact, which in our view is crucial for building trust and for working across cultures. However, more than half of scientists questioned in a European Commission survey (www.more-2.eu) considered that virtual mobility would make short-term visits unnecessary.

Meeting participants agreed that virtual mobility would provide equal access to and for researchers with physical disabilities, would help those on parental leave to maintain contact with their national and international networks, and would enable researchers in poorer regions to access well-resourced labs and to collaborate internationally.

We maintain that virtual mobility should be considered on the same footing as mobility between disciplines, sectors and geographical regions, and that it should be seen as a driver of equal opportunities. Peer review and evaluation structures need to acknowledge these new mobility concepts.

Conor O'Carroll* *European Research Area Steering Group on Human Resources and Mobility, Newry, Northern Ireland.*

ocarroll.conor@gmail.com

**On behalf of 9 correspondents (see go.nature.com/emgqqf for full list).*

China is closing its rural education gap

Schemes are already under way to address the education gap between China's urban and rural areas (see Q. Wang *Nature* **510**, 445; 2014). These are improving education opportunities for rural students and supplying them with the best teachers.

For example, over the past ten years, the popular Go West programme has supplied more than 160,000 leading graduates to support the development of poor rural areas. The 17,500 positions provided this year cover several aspects, including teaching (see go.nature.com/lrhhb6p; in Chinese). The Postgraduate Group of Volunteers to Support Education recruits teachers for rural regions from China's most prestigious universities, such as the Harbin Institute of Technology, and the non-governmental initiative Project Hope is helping to educate poor students.

More people from underdeveloped regions are enrolling in the country's leading universities, thanks to 185,000 government places allocated to students from these areas this year. And some non-governmental organizations are contributing to educational institutions in deprived countryside areas; these include Our Free Sky, which provides teachers.

These efforts to close the education gap will be complemented by the gathering momentum of rural migration to China's cities (see X. Bai *et al.* *Nature* **509**, 158–160; 2014).
Xin Miao *Harbin Institute of Technology, Harbin, China.*
xin.miao@aliyun.com
Christina W. Y. Wong *Hong Kong Polytechnic University, Kowloon, Hong Kong, China.*

Brain project leaders need an open mind

As neuroscientists in Europe who care about the success of research projects large and small in our field, we are dismayed by the publicly reported attitude of the leaders of the Human Brain Project (HBP) towards scientists who have expressed widely supported criticisms of the project in an open letter (<http://neurofuture.eu>; see also *Nature* **511**, 125 and 133–134; 2014).

Instead of acknowledging that there is a problem and genuinely seeking to address scientists' concerns, the project leaders seem to be of the opinion that the letter's 580 signatories are misguided.

The explicit supposition of the HBP leaders that some aspects of neuroscience research could be done in a different way than in the past deserves respect. However, mindful of the sincerity of a number of the well-regarded neuroscientists who have signed the letter as of 11 July, we submit that the likelihood of all 500+ being misguided is remote.

A more enquiring and open-minded attitude to the concerns expressed may prove to be in the best interests of both the information-technology and neuroscience communities.

Richard Morris* *University of Edinburgh, UK.*

r.g.m.morris@ed.ac.uk

**On behalf of 6 correspondents (see go.nature.com/8nmmdu for full list).*

CONTRIBUTIONS

Correspondence may be submitted to correspondence@nature.com after consulting the guidelines at <http://go.nature.com/cmchno>. Alternatively, readers may comment online: www.nature.com/nature.

HIGH-PRESSURE PHYSICS

Piling on the pressure

The machine that houses the world's largest laser, and which stands in for the starship *Enterprise*'s warp core in the film *Star Trek Into Darkness*, has compressed diamond to the density of lead. [SEE LETTER P.330](#)

CHRIS J. PICKARD & RICHARD J. NEEDS

The stars and planets we can see in the night sky were formed by strong gravitational forces that crushed their constituent atoms tightly together at immense pressures. How, on Earth, can we figure out what effect this force has had on the inside of these distant and inaccessible objects? We are confident about the physics that operates in these stars and planets, but what about their chemistry? Predictions abound, but hard experimental data are desperately needed. On page 330 of this issue, Smith *et al.*¹ present the results of groundbreaking experiments on the compression of carbon diamond up to a pressure similar to that at the centre of Saturn.

The machine used to perform the experiments, the US National Ignition Facility (NIF), is unique (Fig. 1). It houses the world's largest laser, which can be focused onto a millimetre-scale target held at the centre of a 10-metre aluminium sphere. It certainly looks the part: indeed, it stood in for the starship *Enterprise*'s warp core in the movie *Star Trek Into Darkness*. The NIF's primary mission is to study inertially confined nuclear fusion², but a portion of the laser 'shots' have been allocated to fundamental science — from laboratory astrophysics to plasma physics and planetary science.

The new NIF experiments have succeeded in compressing diamond up to a pressure of 5 terapascals (5×10^{12} Pa) — 14 times the pressure at the centre of the Earth. In addition to their brute power, the laser pulses can be exquisitely manipulated, allowing the pressure in the target to be increased in a precisely controlled manner known as dynamic ramped compression. Dynamic compression can generate enormous pressures far beyond those accessible in static experiments that use, for example, diamond anvil cells³. A crucial aspect of the current set-up is that the use of ramped compression reduces the dissipative heating of the sample. Ramped compressions can explore materials at conditions similar to those encountered deep within large planets, whereas compressions using shock waves generally lead to higher temperatures.

The discovery of multiple planets beyond our Solar System, many of which are much

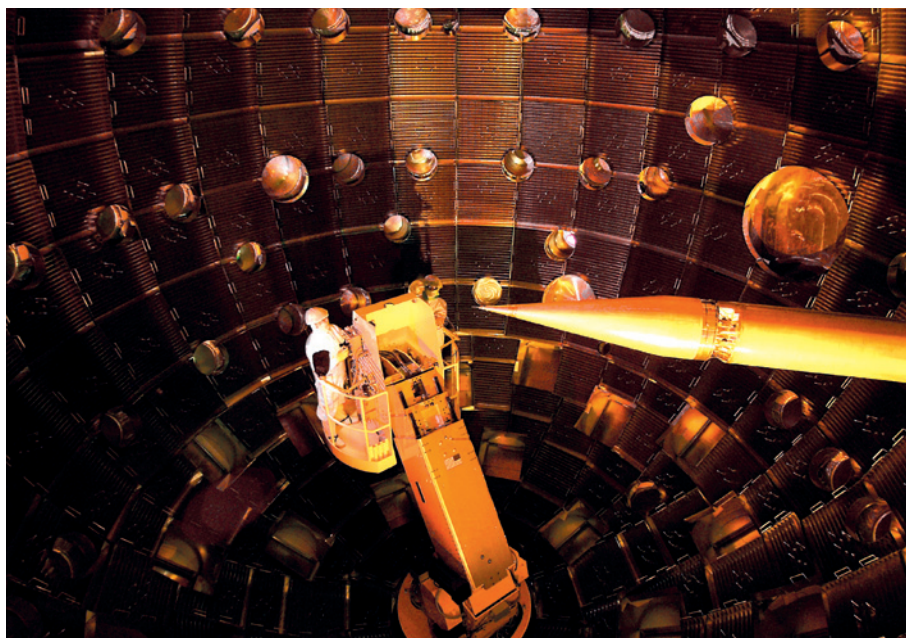


Figure 1 | The National Ignition Facility machine. Smith *et al.*¹ have used this machine to compress diamond up to a record pressure of 5×10^{12} pascals.

larger than Jupiter and Saturn, has led to a dramatic change in our picture of the Universe. Understanding the make-up and evolution of these exoplanets requires the development of theoretical models, which depend on the pressure–density equations of state of the most likely planetary materials⁴. Until now, these equations of state have largely been determined by extrapolating from terrestrial data.

Extrapolation is a perilous activity. Theoretical calculations of terapascal-pressure phase transitions in, for example, aluminium (which is used in high-pressure dynamical experiments as a standard material with well-understood properties) predict that it will transform from a close-packed structure to a complicated non-close-packed structure at terapascal pressures. At pressures higher than those investigated by the authors, some of the valence electrons of carbon are expected to move away from the nucleus and play the part of the fluorine anions in ionic calcium fluoride, with the calcium sites being occupied by carbon cations⁵. All of this suggests that the structures adopted at terapascal pressures

may be surprising and far from simple⁶.

Simple quantum-mechanical theories, such as the Thomas–Fermi–Dirac theory for very hot dense matter, and more sophisticated quantum algorithms, including both the path-integral Monte Carlo method for ‘warm’ dense matter and density functional theory for condensed phases, have been shown to provide largely consistent descriptions in the pressure–density regions where their applicability overlaps⁷. For the pressures and densities probed by the current experiments, a series of phase transitions is predicted to occur in which carbon becomes denser than in its diamond form. Interestingly, the experiments did not detect any of these phase transitions, which may have been smoothed out, or deferred, through some as-yet-unknown mechanism. Overall, however, the agreement between results from density-functional-theory calculations and the experiment is good, so the theory is likely to be on a solid footing.

The authors are confident that their carefully designed dynamic ramped compression has achieved temperatures that are similar to those inside planets. Although the temperatures

LAWRENCE LIVERMORE NAT'L LAB

generated in their experiments can be inferred through theoretical predictions, Smith *et al.* cannot directly measure the actual temperatures. In addition, it is not currently possible to use their methods to determine crystal structures at terapascal pressures. These are exciting challenges for the future. Important progress has been made in this direction⁸, and there is hope that laser-driven dynamic compression, coupled with free-electron lasers, will provide diagnostic snapshots of structures and their dynamics.

Planets form over many millions of years, whereas the reported dynamic ramped compression procedure is over in a flash. It is not clear whether these experiments, despite reaching relevant temperatures and pressures, are able to closely model the largely equilibrated, dense rocks and ices existing within giant planets. However, the brevity of the experiments does have an advantage. Just as nanotechnology has been a gift to theoreticians, allowing meaningful computations of manageable numbers of atoms, the short experimental timescales actually make the behaviour of compressed atoms easier to model in dynamical simulations. Through mutual benchmarking and the testing of predictions, we expect that experiment and theory will together improve our understanding of matter under extreme compression.

A final note of perspective. Although the pressures and densities probed in the current experiments are immense, nature is even more ambitious. The giant exoplanets are a stepping stone to the stars, where petapascal pressures (1 petapascal is 10^{15} Pa) are reached. The predictions of rich terapascal-pressure physics should caution against assumptions of simple structures. Indeed, a recent theoretical study⁹ anticipates a complex metallurgy for the crusts of neutron stars. Over to the experimenters! ■

Chris J. Pickard is in the Department of Physics and Astronomy, University College London, London WC1E 6BT, UK, and at the London Institute for Mathematical Sciences. **Richard J. Needs** is in the Theory of Condensed Matter Group, Cavendish Laboratory, Cambridge CB3 0HE, UK. e-mail: c.pickard@ucl.ac.uk

1. Smith, R. F. *et al.* *Nature* **511**, 330–333 (2014).
2. Ball, P. *Nature* <http://dx.doi.org/10.1038/nature.2014.14710> (2014).
3. Jeanloz, R. *et al.* *Proc. Natl Acad. Sci. USA* **104**, 9172–9177 (2006).
4. Wagner, F. W., Sohl, F., Hussmann, H., Grott, M. & Rauer, H. *Icarus* **214**, 366–376 (2011).
5. Martinez-Canales, M., Pickard, C. J. & Needs, R. J. *Phys. Rev. Lett.* **108**, 045704 (2012).
6. Pickard, C. J. & Needs, R. J. *Nature Mater.* **9**, 624 (2010).
7. Driver, K. P. & Militzer, B. *Phys. Rev. Lett.* **108**, 115502 (2012).
8. Rygg, J. R. *et al.* *Rev. Sci. Instrum.* **83**, 113904 (2012).
9. Kobayakov, D. & Pethick, C. J. *Phys. Rev. Lett.* **112**, 112504 (2014).

ECOLOGY

Pesticides linked to bird declines

Decreases in bird numbers are most rapid in areas that are most heavily polluted with neonicotinoids, suggesting that the environmental damage inflicted by these insecticides may be much broader than previously thought. [SEE LETTER P.341](#)

DAVE GOULSON

The debate over the environmental risks posed by neonicotinoid insecticides has raged since the late 1990s, when French beekeepers began blaming the chemicals for losses of honeybee colonies. The discussion has focused closely on bees, particularly the risks posed by the use of neonicotinoid treatments on flowering crops that bees visit. But on page 341 of this issue, Hallmann *et al.*¹ provide strong evidence that this debate may have missed the bigger picture. Analysing long-term data sets on bird populations in the Netherlands, the authors demonstrate that regional patterns of population decline in insect-eating birds are neatly predicted by levels of neonicotinoids detected in environmental samples. In other words, birds have declined faster in places with more neonicotinoid pollution.

Dozens of papers have been published on the effects of neonicotinoids on bees and, following a review of the evidence, the European Food Safety Authority declared in 2013 that neonicotinoids posed an “unacceptable risk” to the insects. Shortly afterwards, the European Union voted in favour of a two-year moratorium on the use of three widely used neonicotinoids on flowering crops. It has

already been suggested that the impacts of these chemicals are likely to extend far beyond bees², but Hallmann and colleagues’ study is the first to provide direct evidence that the widespread depletion of insect populations by neonicotinoids has knock-on effects on vertebrates.

Neonicotinoids are neurotoxins that are exceptionally toxic to insects but much less so to birds³. Because of this, the observed bird declines are unlikely to be due to direct toxicity. As Hallmann *et al.* argue, it is much more plausible that the effects are the result of a depletion of the birds’ food — insects. However, it is worth noting that none of the bird species studied would ordinarily eat bees in any quantity.

Hallmann and colleagues essentially infer cause and effect from correlation, but this is made more convincing because they consider a range of other measures of land use that are known to affect bird and insect populations, but found none that predicted bird declines as powerfully as environmental neonicotinoid concentration. Of course, an experimental, manipulative approach to test cause and effect would be more compelling, but that would be almost impossible on a realistic scale, with replication, in organisms as highly mobile as birds, and in any case would face severe ethical issues.

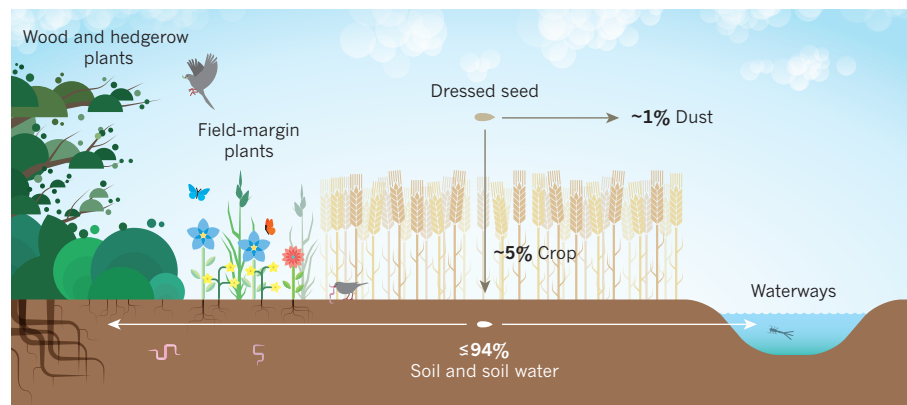


Figure 1 | The environmental fate of neonicotinoids. When neonicotinoids are applied as a seed dressing to crops, the bulk of the active ingredients (80–98%) enter the soil and soil water. There, they can persist for long periods, accumulate, be taken up by the roots of vegetation at the margins of fields and follow-on crops, or leach into aquatic systems. Neonicotinoids are highly toxic to insects, which are exposed to the chemicals in plants, soil and water. Hallmann *et al.*¹ have observed rapid declines in bird populations in regions with high environmental neonicotinoid concentrations, and suggest that they are the result of insect poisoning depleting the birds’ food supply.

How might neonicotinoids, most of which are applied as seed dressings to arable crops, come to have such widespread impacts on the environment? The insecticides' intended mechanism of action is that the dressing should dissolve around the seed, be absorbed by the growing seedling and spread through its tissues, protecting all parts of the crop from herbivorous insects. However, only approximately 5% of the active ingredient is taken up by the crop⁴ (Fig. 1). A little is lost as toxic dust that blows away and may affect flying insects or be deposited on non-target vegetation⁵, but most enters the soil and soil water. The half-life of neonicotinoids in soil varies with soil type, but can exceed 1,000 days, such that they can accumulate over time. The consequences of this accumulation for soil fauna and soil health are poorly understood².

The chemicals can also be washed from soils into waterways, where they are likely to affect aquatic insects⁶, which are key sources of food for both birds and fish. And they can be taken up by the roots of hedgerow plants, where they will have the same systemic action as in crops, spreading through the leaves and flowers. Non-target herbivorous insects such as grasshoppers, beetles, shield bugs and the caterpillars of butterflies, moths and sawflies will all be exposed through this route, and these form the food supply for a broad range of predatory insects, birds and some mammals, such as shrews and bats.

The persistent nature of neonicotinoids and their high solubility in water mean that such broad contamination is also probable with other methods of application, such as foliar sprays or soil drenches. Given these manifold routes of spread, it is perhaps not surprising that, after 20 years of steadily increasing use, there is now evidence that neonicotinoids are having broad effects through the food chain — as shown by Hallmann *et al.* and by a recent meta-analysis⁷ of studies on the ecosystem effects of systemic pesticides.

The European two-year moratorium came into effect in December 2013, but it is designed to protect bees from exposure only to mass-flowering crops. As such, neonicotinoids are still used as seed dressings on other major crops, such as wheat and barley, and they are still widely sprayed in horticulture and sold for use in gardens and public areas. Hence, impacts on birds and other insectivores might be expected to continue. Elsewhere in the world, the emerging evidence for environmental harm has not yet resulted in any new restrictions on their use.

The story is reminiscent of Rachel Carson's *Silent Spring*⁸, published in 1962. She wrote: "These sprays, dusts, and aerosols are now applied almost universally to farms, gardens, forests, and homes — nonselective chemicals that have the power to kill every insect, the 'good' and the 'bad', to still the song of birds and the leaping of fish in the streams..."

Carson was describing the environmental devastation caused by the over-reliance on and overuse of organochloride insecticides such as DDT (dichlorodiphenyltrichloroethane) in the 1950s and 1960s, which led to major problems with outbreaks of pesticide-resistant pests, widespread contamination of the environment and knock-on effects through the food chain, including chronic poisoning of people. She would undoubtedly think that we seem to have learnt little from our past mistakes. ■

Dave Goulson is in the School of Life Sciences, Sussex University, Falmer BN1 9QG, UK.
e-mail: d.goulson@sussex.ac.uk

ASTROPHYSICS

Survival of the largest

Whether supernovae create most of the dust in the cosmos is a controversial question. Observations of a distant supernova have revealed signs of freshly formed dust, but the properties of the dust are unexpected. SEE LETTER P.326

HALEY GOMEZ

Dust grains play a crucial part in galaxy evolution. They aid in the formation of stars and provide the building blocks of rocky planets and life itself. However, the origin of dust is a contentious topic: it remains unclear whether dust is formed in the violent deaths of massive stars. Supernova explosions are often portrayed as the villains in the life cycle of dust in galaxies, with the harsh million-kelvin gas of the debris thought to efficiently destroy dust grains — produced by the supernova and in the surrounding material — through high-speed collisions with atoms and other grains^{1,2}. But indirect observations of considerable quantities of dust in galaxies at low and high redshifts suggest either that supernovae are producing lots of dust^{3–6} or that dust destruction by the supernovae is inefficient. In this issue, Gall *et al.*⁷ (page 326) describe observations of telltale signatures from dust in an extragalactic supernova. The results reveal, for the first time, that both of these scenarios are likely to be true.

Over the past few years, thanks in part to far-infrared, millimetre and submillimetre telescopes such as the European Space Agency's Herschel Space Observatory^{8,9} and the Atacama Large Millimeter/submillimeter Array (ALMA)^{10,11}, evidence has slowly mounted that dust formation in the aftermath of a supernova may in fact be ubiquitous¹². In their study, Gall and colleagues investigated whether dust grains were present in the distant supernova 2010jl (Fig. 1). They did this by checking for signs of absorption of light — owing to dust within the supernova — from

Twitter: @DaveGoulson

1. Hallmann, C. A., Foppen, R. P. B., van Turnhout, C. A. M., de Kroon, H. & Jongejans, E. *Nature* **511**, 341–343 (2014).
2. Goulson, D. *J. Appl. Ecol.* **50**, 977–987 (2013).
3. Tomizawa, M. & Casida, J. E. *Annu. Rev. Pharmacol. Toxicol.* **45**, 247–268 (2005).
4. Sur, R. & Stork, A. *Bull. Insectol.* **56**, 35–40 (2003).
5. Tapparo, A. *et al. Environ. Sci. Technol.* **46**, 2592–2599 (2012).
6. Van Dijk, T. C., Van Staalduinen, M. A. & Van der Sluijs, J. P. *PLoS ONE* **8**, e62374 (2013).
7. van der Sluijs, J. P. *et al. Environ. Sci. Pollut. Res.* <http://dx.doi.org/10.1007/s11356-014-3229-5> (in the press).
8. Carson, R. *Silent Spring* (Houghton Mifflin, 1962).

This article was published online on 9 July 2014.

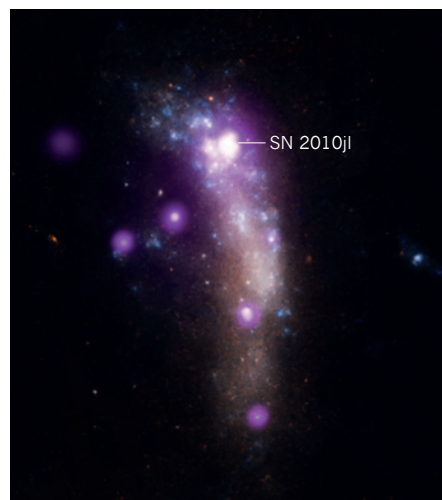


Figure 1 | Supernova explosion in a distant galaxy. Gall and colleagues⁷ examined the dust content of supernova 2010jl, which exploded in a galaxy about 50 million parsecs away from Earth. The system is seen here in an image that combines X-ray and optical observations. The image is about 46 arcseconds across.

debris moving towards and away from us, and by searching for thermal emission from dust in the near-infrared (NIR) part of the electromagnetic spectrum.

Using the Very Large Telescope in Chile, the team observed the supernova over 10 epochs starting 26 days after the initial explosion, and found clear evidence that dust grains were formed in the dense shell that lies just behind the expanding supernova shock. They found that, by day 868 after the explosion, the amount of dust in the supernova had grown considerably compared with their observations at earlier

X-RAY: NASA/CXC/R. MILIT. COLL. CANADA/P. CHANDRA ET AL.; OPTICAL: NASA/STSC

epochs. From the NIR emission, they derived a mass of dust equivalent to 830 Earth masses, which is 40 times lower than observed in the ancient Crab Nebula supernova remnant⁹. Such a small dust mass is unsurprising, given the relative youthfulness of SN 2010jl.

As well as measuring the quantity of freshly formed dust, Gall *et al.* used their data to graphically show the extent of the absorption of light by the dust grains as a function of wavelength — the extinction curve. This curve provides information on the dust composition (carbon-rich in this case) and size distribution, and reveals perhaps the most significant result from this work: newly formed supernova dust grains are gigantic compared with dust typically found in our Galaxy. The same type of analysis for the Milky Way requires dust grains with a maximum size of 0.25 micrometres to reproduce the observed extinction curve, but in SN 2010jl, the grains need to be greater than 1 µm with a maximum grain radius of 4.2 µm.

The presence of such large grains in a distant supernova is at odds with the size distribution assumed in theoretical dust models used in the literature¹³. However, this is not the first time that astronomers have observed large grains. The Ulysses robotic spacecraft mission¹⁴ recorded substantial emission from grains larger than 2 µm entering our Solar System, and grains as large as 6 µm were detected hitting our planet's atmosphere¹⁵. Similarly large dust grains have also been seen in distant γ-ray bursts¹⁶.

These large grains seen in our Solar System, and now in an extragalactic supernova, imply not only that dust is created directly as a result of the explosion, but also that supernova dust might be hardy enough to survive the explosion's harsh environment. Owing to their size, larger grains will be more resilient to high-speed collisions compared with smaller grains, and could well survive the explosion in the long term, albeit chipped into smaller pieces as they make their way into the surrounding gas.

Another supernova (SN 1987A) in the nearby Large Magellanic Cloud, a satellite galaxy of the Milky Way, perhaps provides researchers with an ideal laboratory to directly measure the efficiency of dust destruction in supernova shocks. The debris of SN 1987A^{10,11} is currently moving at 2,000 kilometres per second, and will soon collide with a ring of material left over from the progenitor star before the explosion. Astronomers will be able to observe with ALMA the thermal emission from the dust as the supernova ejecta and the ring collide in real time. Such observations will detect an evolution in dust formation and destruction even at a distance of 50,000 parsecs (the distance from Earth at which the debris of SN 1987A is located).

► NATURE.COM
For more on the origins of dust in the Universe, visit:
go.nature.com/hovgol

If collisions do prove to be less destructive than theoretical models currently suggest, this will be comforting news to

astronomers trying to explain the large dust masses observed in galaxies^{6,7,17}. It seems that supernovae may not be the bad guys after all. ■

**Haley Gomez is in the School of Physics and Astronomy, Cardiff University, Cardiff CF24 3AA, UK.
e-mail: haley.gomez@astro.cf.ac.uk**

1. Jones, A. P., Tielens, A. G. G. M., Hollenbach, D. J. & McKee, C. F. *Astrophys. J.* **433**, 797–810 (1994).
2. Barlow, M. & Silk, J. *Astrophys. J.* **211**, L83–L87 (1977).
3. Gall, C., Hjorth, J. & Andersen, A. C. *Astron. Astrophys. Rev.* **19**, 43 (2011).
4. Morgan, H. L. & Edmunds, M. G. *Mon. Not. R. Astron. Soc.* **343**, 427–442 (2003).
5. Matsuura, M. *et al.* *Mon. Not. R. Astron. Soc.* **396**, 918–934 (2009).

6. Dunne, L. *et al.* *Mon. Not. R. Astron. Soc.* **417**, 1510–1533 (2011).
7. Gall, C. *et al.* *Nature* **511**, 326–329 (2014).
8. Pilbratt, G. L. *et al.* *Astron. Astrophys.* **518**, L1 (2010).
9. Gomez, H. L. *et al.* *Astrophys. J.* **760**, 96–108 (2012).
10. Matsuura, M. *et al.* *Science* **333**, 1258–1261 (2011).
11. Indebetouw, R. *et al.* *Astrophys. J. Lett.* **782**, L2 (2014).
12. Gomez, H. L. *Proc. Sci.* http://pos.sissa.it/archive/conferences/207/146/LCDU2013_146.pdf (2014).
13. Zubko, V., Dwek, E. & Arendt, R. G. *Astrophys. J. Suppl.* **152**, 211–249 (2004).
14. Grün, E. *et al.* *Nature* **362**, 428–430 (1993).
15. Meisel, D. D., Janches, D. & Mathews, J. D. *Astrophys. J.* **579**, 895–904 (2002).
16. Li, Y., Li, A. & Wei, D. M. *Astrophys. J.* **678**, 1136–1141 (2008).
17. Rowlands, K. *et al.* *Mon. Not. R. Astron. Soc.* **441**, 1040–1058 (2014).

This article was published online on 9 July 2014.

NEUROBIOLOGY

Keeping a lid on it

The protein Npas4 dampens activated excitatory brain circuits by recruiting inhibitory signals to excitatory neurons. It emerges that this protein has the opposite role in some inhibitory neurons, promoting their activity.

GINA TURRIGIANO

The astounding abilities of the mammalian brain arise from a few core circuit ‘motifs’. One such motif is positive feedback¹, in which the mutual excitation of pyramidal neurons amplifies small signals. Now, fans of rock legend Jimi Hendrix will immediately recognize the problem this raises: positive-feedback amplification can easily get out of control, and an effect that is awesome in ‘Voodoo Child’ can lead to epilepsy in brain circuits. Our brains must therefore counteract positive feedback with inhibitory circuit motifs — pyramidal neurons excite several subtypes

of inhibitory neuron, which then inhibit those same pyramidal neurons through negative feedback (Fig. 1a). One mystery is how these circuits are adjusted to maintain the excitation–inhibition balance in the brain². Writing in *Cell*, Spiegel *et al.*³ provide insight into this homeostatic balancing act, showing how gene-expression pathways that regulate neuronal circuits are differentially tuned to the function of inhibitory and excitatory motifs.

During development, neuronal identity is determined by the restriction of gene expression to a subtype-specific pattern⁴. However, gene expression does not then remain static. For our brains to learn and adapt, neurons

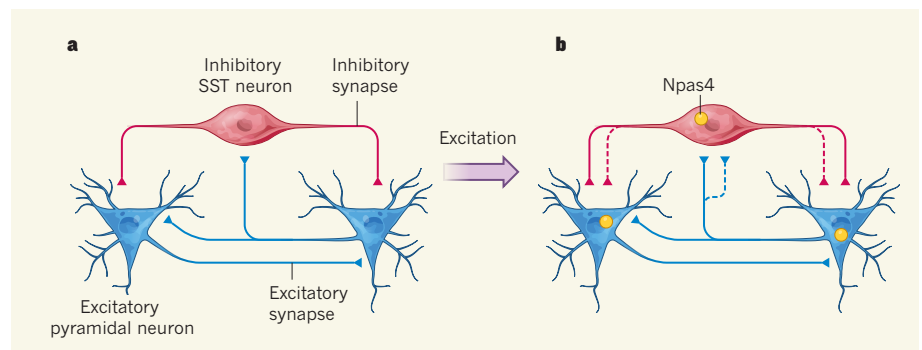


Figure 1 | Balancing excitation levels. **a**, Excitatory pyramidal neurons transmit signals to inhibitory somatostatin-positive (SST) neurons, and vice versa, through neurotransmitting junctions called synapses. In addition, excitatory neurons synapse to one another in a positive feedback loop. **b**, Spiegel *et al.*³ report that neural excitation induces expression of the transcription factor Npas4 in both cell types, triggering neuron-specific gene programs. Npas4 expression in SST neurons causes an increase in the number of excitatory synapses to these neurons (blue dashed synapse). Conversely, Npas4 expression in pyramidal neurons increases their inhibition (red dashed synapses). Overall, these dynamic changes dampen excitation.

must respond to changes in the environment, and this dynamism arises in part through activity-dependent changes in gene-expression pathways⁵. These changes are thought to control activity by, for example, adjusting the effectiveness of excitatory and inhibitory synaptic connections (junctions between neurons that transmit information) in a manner that is specific to both cell and synapse type⁶. For instance, too much activity boosts the effectiveness of inhibitory synapses acting on excitatory neurons, dampening excitation. Conversely, too little activity increases the effectiveness of excitatory synapses acting on excitatory neurons. Thus, homeostatic plasticity follows a 'circuit logic' that coordinately adjusts excitatory and inhibitory feedback loops to stabilize neuron firing⁶.

Spiegel and colleagues set out to identify genes that contribute to such neuronal-subtype-specific adjustments. To do this, they generated neuronal cultures that were enriched in either inhibitory or excitatory neurons. When the authors depolarized the cultures (which mimics excitation), the two cell types displayed similar early changes in gene expression. In particular, the expression of several early-response genes, including *Npas4*, was increased in both cultures.

Things got interesting when Spiegel and co-workers turned their attention to the late response to depolarization. After six hours, there was a substantial increase in the number of genes whose expression was modified, but the fraction of modified genes that was shared by inhibitory and excitatory neurons was smaller than during the early response. The authors then confirmed these results *in vivo* using an approach that allowed them to probe gene expression in a cell-type-specific manner. Taken together, their results suggest that enhanced activity triggers a shared early transcriptional program in excitatory and inhibitory neurons, which then sets in motion distinct downstream signalling pathways.

The early-response gene *Npas4* caught Spiegel and colleagues' attention because the transcription factor that it encodes⁷ acts to promote homeostasis in excitatory pyramidal neurons by regulating the number of inhibitory synapses they receive⁸. The authors wondered whether *Npas4* might have a different function in inhibitory neurons, because enhancing inhibition onto inhibitory neurons would have the paradoxical effect of activating pyramidal neurons — a counterproductive effect for homeostasis.

To test this, Spiegel *et al.* manipulated *Npas4* expression in somatostatin-positive (SST) inhibitory neurons, which mediate a type of feedback inhibition in the brain. Selectively removing *Npas4* from SST neurons in brain slices or in cultures containing both inhibitory and excitatory cell types had no effect on the number of inhibitory synapses to SST neurons, but decreased excitatory synapses.

Conversely, overexpressing *Npas4* in SST neurons increased excitatory synapses to those neurons. Furthermore, the authors found that *Npas4* deletion compromised the expression of a subset of late-response genes in SST neurons, but that *Npas4* overexpression promoted expression of these same genes.

Spiegel *et al.* therefore conclude that enhanced neuronal activity activates *Npas4* in both cell types. This sets in motion different late-response transcriptional programs that have distinct outcomes — increased excitation of SST neurons and increased inhibition of pyramidal neurons. These two *Npas4*-mediated gene programs would be expected to synergize, overall inducing increased inhibition of pyramidal neurons and thus counteracting a rise in activity (Fig. 1b).

Although the model is appealing, it is important to bear in mind that brain circuits contain several subtypes of inhibitory neuron, and that the SST-pyramidal circuit is only one of many feedback loops that regulate excitability¹. Whether the changes measured here contribute significantly to circuit homeostasis remains unknown.

A second caveat is that, although the model predicts that raising activity should increase excitatory synapses to SST neurons in an *Npas4*-dependent manner, Spiegel and colleagues did not test this prediction directly. Despite the fact that directly reducing or increasing *Npas4* expression does modulate synapse number, the effects of *Npas4* when manipulated alone may be different from its effects in the context of other activity-induced genes. As such, experiments

that confirm the authors' model seem key.

Finally, the study raises the fundamental question of how *Npas4* regulates distinct genes in different cell types. Spiegel *et al.* find a partial answer — regulatory DNA elements that control the expression of *Npas4* target genes are in different epigenetic states in the two cell types (epigenetic regulation changes gene expression without altering DNA sequence). This suggests that gene programs underlying homeostasis are epigenetically tuned to the function of each neuron within a neural circuit.

So if listening to Hendrix amps your brain circuits up to 11, don't worry. Dynamic negative feedback loops, working through cell-type-specific effectors such as *Npas4*, are there to keep a lid on things. ■

Gina Turrigiano is in the Department of Biology, Center for Complex Systems, and at the Center for Behavioral Genomics, Brandeis University, Waltham, Massachusetts 02454, USA.

e-mail: turrigiano@brandeis.edu

1. Hangya, B., Pi, H.-J., Kvitsiani, D., Ranade, S. P. & Kepecs, A. *Curr. Opin. Neurobiol.* **26**, 117–124 (2014).
2. Turrigiano, G. G. & Nelson, S. B. *Nature Rev. Neurosci.* **5**, 97–107 (2004).
3. Spiegel, I. *et al.* *Cell* **157**, 1216–1229 (2014).
4. Edlund, T. & Jessell, T. M. *Cell* **96**, 211–224 (1999).
5. Lyons, M. R. & West, A. E. *Prog. Neurobiol.* **94**, 259–295 (2011).
6. Turrigiano, G. *Cold Spring Harb. Perspect. Biol.* **4**, a005736 (2012).
7. Maya-Vetencourt, J. F. *Neural Plast.* **2013**, 683909 (2013).
8. Lin, Y. *et al.* *Nature* **455**, 1198–1204 (2008).

This article was published online on 9 July 2014.

CANCER

Sugar-coated cell signalling

Cell membranes are covered with sugar-conjugated proteins. New findings suggest that the physical properties of this coating, which is more pronounced in cancer cells, regulate cell survival during tumour spread. [SEE ARTICLE P.319](#)

ANDREW J. EWALD & MIKALA EGEBLAD

The cell membrane serves as a signalling interface that allows cells to exchange information with their environment. It is constructed from lipids and contains both transmembrane and lipid-tethered proteins, which can be further modified through the covalent addition of sugars to build glycoproteins. Cancer cells frequently have higher levels of glycoproteins, such as mucin-1 (refs 1–3), than do healthy cells, and individual glycoproteins can transduce environmental

signals that directly promote malignancy. However, glycoproteins also collectively organize into a glycocalyx. In this issue, Paszek *et al.*⁴ (page 319) show how the physical properties of this coating regulate the clustering of cell-surface receptors and thereby affect intracellular signalling in ways that can contribute to cancer metastasis.

The authors demonstrate that the thickness of the glycocalyx is a crucial determinant of the spatial and temporal features of receptor–ligand interactions. Specifically, they find that the thick glycocalyx of cancer cells serves as

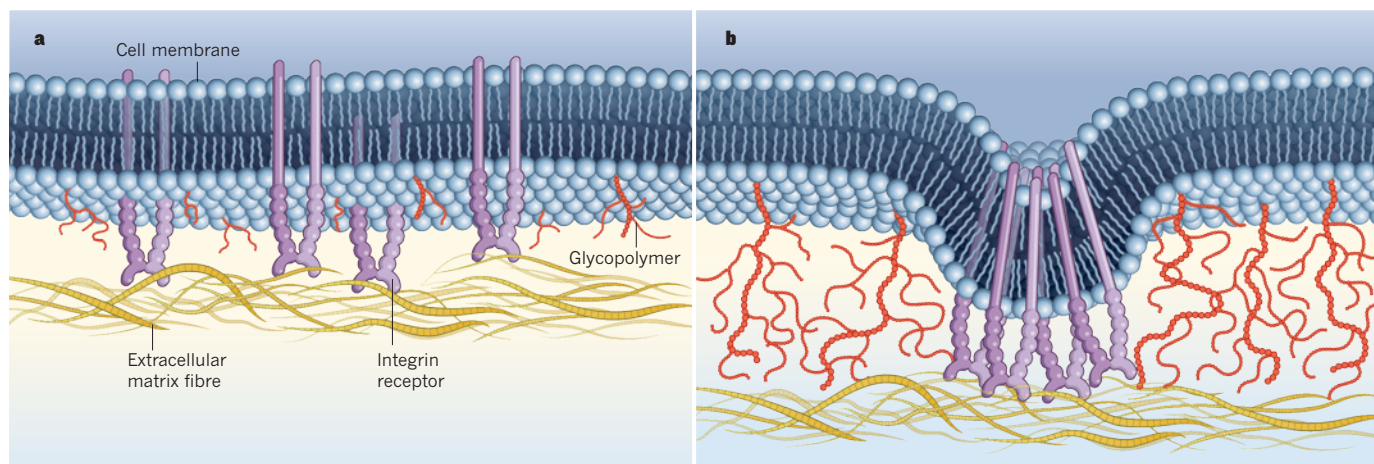


Figure 1 | Clustering for survival. **a**, Paszek *et al.*⁴ show that cells with short synthetic glycopolymers (which mimic the physical properties of glycoproteins) attached to their cell membrane exhibit a close gap between the membrane and the extracellular matrix (ECM) and a relatively uniform distribution of glycopolymers and integrins in the membrane. **b**, By contrast, the presence of long synthetic glycopolymers or the natural glycoprotein mucin-1 (not shown) results in an expanded membrane–ECM gap, clustering of integrins, the exclusion of glycopolymers from regions of integrin adhesion, and membrane bending. These physical effects alter cell signalling through the MEK, PI3K and FAK pathways, leading to enhanced cell survival.

a ‘kinetic trap’, generating regions on the cell surface where the likelihood of receptor–ligand interactions is increased, driving receptor clustering (Fig. 1). Integrins are transmembrane receptors that bind extracellular matrix (ECM) proteins and are key interpreters and integrators of both the biochemical composition and the mechanical properties of the extracellular space^{5,6}. Paszek and colleagues reveal that cells with a thick glycocalyx are more efficient at receiving cell-survival signals through integrins, owing to the kinetic-trap properties of the glycocalyx. This may facilitate metastatic spread by enabling cancer cells to survive in the varied tissue and fluid environments they must traverse to colonize distant organs.

To uncouple the signalling properties of individual glycoproteins from the more general consequences of a bulky glycocalyx, the authors generated a series of synthetic glycopolymers to mimic the physical properties of glycoproteins of different sizes. They then tested how glycopolymers that projected 3 nanometres, 30 nm or 80 nm into the extracellular space influenced signalling through integrins, which have a reported length⁷ of about 20 nm. The long (80 nm) glycopolymers expanded the average gap between the cell membrane and the extracellular matrix and, as predicted by previous computational modelling⁸, reduced the overall rate of integrin binding to the ECM. New integrin–ECM interactions occurred preferentially near existing adhesion sites, thereby increasing the focal clustering of integrins on the cell surface (Fig. 1). The long glycopolymers were excluded from these clusters. By contrast, the short and medium-length synthetic glycopolymers did not affect integrin clustering, even when present at high surface densities.

The authors next evaluated the effects of the natural glycoprotein mucin-1 (Muc1), which is 10–100-fold upregulated in many cancers^{1–3} and extends 200 nm or more from the cell

surface. Like the long synthetic glycopolymers, Muc1 expression increased the cell–ECM gap, increased total cell–ECM adhesion and enhanced the size of integrin clusters. As predicted by the kinetic-trap model, ECM-ligated integrins rarely entered regions occupied by Muc1. None of these effects required the signalling-competent cytoplasmic tail of Muc1, revealing a key role for the physical properties of the extracellular part of the glycoprotein.

Integrin-based cell–matrix signalling is important for many steps in metastasis, including the migration of cancer cells out of the primary tumour and through the ECM, their entry into the vasculature, survival in the circulation, adhesion to the vessel wall, exit from the vasculature, and migration to and proliferative expansion in a distant organ⁶. By reducing the rate of integrin binding and promoting clustering at existing adhesion sites, bulky glycoproteins act to promote a stable interaction between the cancer cells and the ECM.

Such stability is probably not optimal for the turnover of adhesions that is necessary for rapid migration. However, for cancer cells to metastasize, they must not only disseminate from the primary tumour to the secondary organ, but also survive in the many different microenvironments that they travel through. Integrins play a major part in cell survival⁹, as well as cell migration. Normal cells initiate a process of programmed cell death when they lack appropriate integrin ligation, and ECM–integrin binding therefore represents a mechanism for keeping cells in the correct place in the body. Paszek *et al.* demonstrate that bulky glycoproteins lower the threshold for reaching sufficient integrin ligation to survive and proliferate. This effect requires signalling through the MEK, PI3K and FAK intracellular pathways. They also show that the cytoplasmic domain of Muc1 is dispensable for its effects on cell survival, supporting the idea that the physical properties of

the glycocalyx influence cell signalling.

This exciting paper establishes a new conceptual framework for the biological function of cell-surface glycoproteins. Independent of, and in addition to, their biochemical properties, the bulky constituents of the glycocalyx physically influence the spatial organization of integrin receptors and hence their activity. These effects are likely to be common to other cell-surface receptors that are regulated by receptor clustering or related intermolecular interactions. It will therefore be interesting to evaluate how the glycocalyx regulates other major signalling pathways. We expect that the optimal glycocalyx thickness for supporting different aspects of cancer-cell behaviour, including invasion, vascular spread and metastatic colonization, varies. But how cancer cells adapt their glycocalyx to the diverse surroundings that they experience during metastasis is an interesting open question. ■

Andrew J. Ewald is in the Departments of Cell Biology and Oncology, School of Medicine, Johns Hopkins University, Baltimore, Maryland 21218, USA. **Mikala Egeblad** is in the Cancer Center, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA.
e-mail: egeblad@cshl.edu

1. Horm, T. M. & Schroeder, J. A. *Cell Adhes. Migrat.* **7**, 187–198 (2013).
2. Hakomori, S. *Cancer Res.* **56**, 5309–5318 (1996).
3. Hollingsworth, M. A. & Swanson, B. J. *Nature Rev. Cancer* **4**, 45–60 (2004).
4. Paszek, M. J. *et al.* *Nature* **511**, 319–325 (2014).
5. Hynes, R. O. & Naba, A. *Cold Spring Harb. Perspect. Biol.* **4**, a004903 (2012).
6. Hood, J. D. & Cheresh, D. A. *Nature Rev. Cancer* **2**, 91–100 (2002).
7. Campbell, I. D. & Humphries, M. J. *Cold Spring Harb. Perspect. Biol.* **3**, a004994 (2011).
8. Paszek, M. J., Boettiger, D., Weaver, V. M. & Hammer, D. A. *PLoS Comput. Biol.* **5**, e1000604 (2009).

This article was published online on 25 June 2014.

A superelastic organic crystal

Superelasticity — a form of elasticity that involves a phase transition — has been observed for the first time in a pure organic crystal. The material could find applications in microfluidics.

TOMIKI IKEDA & TORU UBE

Since its discovery¹ in 1932, in a gold–cadmium alloy, the property of superelasticity has never been observed in organic crystals — until now. Writing in *Angewandte Chemie*, Takamizawa and Miyamoto² report the discovery of this phenomenon in a single crystal of a simple organic molecule, terephthalamide.

In metallic alloys and ceramic materials, the individual components are strongly bound to one another, forming hard crystals. Under an applied stress, some of these materials can undergo a phase transformation, which can lead to macroscopic deformation³. On removal of the stress, the new phase becomes unstable and the initial phase reappears, and with it the original shape. A typical class of such superelastic material is shape-memory alloys³,

which can deform to up to 10% of their original size but return to their pre-deformed shape. Titanium–nickel alloys are the main type of shape-memory materials, and have applications in devices such as medical stents and spectacle frames⁴.

Takamizawa and Miyamoto examined a soft crystal of terephthalamide about 150 micrometres thick and 59 μm wide. The crystal was initially in what is called the α phase (mother phase) and was pushed with a metal blade, 25 μm wide, against one crystal surface at a speed of 500 μm per minute. The authors found that, when the stress applied by the blade reached a constant value, the crystal underwent a phase transformation into a daughter phase (β phase) at the contacting area between the blade and the surface. Interestingly, the daughter phase grew first along the pushing direction of the blade,

but when this phase hit the bottom of the crystal it grew at a right angle to the pushing direction (Fig. 1).

As the daughter-phase region grew and propagated, the crystal bent at the interface between the two phases. When the authors pulled the blade back, the area of the daughter phase started to decrease and the crystal underwent the reverse phase transformation, eventually reverting to its original shape. The researchers repeated this transformation cycle 100 times, and showed that the crystal deformed by up to 11.34% of its original shape. The stress necessary to induce the transformation from the mother to the daughter phase is roughly 1,000 times smaller than that for the equivalent transformation in the typical titanium–nickel alloy.

In the terephthalamide crystal, molecules associate to form sheets that are held together by a network of hydrogen bonds. Because each terephthalamide molecule has four sites with which to form hydrogen bonds, owing to the presence of an amide group (CONH_2) at each end of the molecule's benzene ring, the network contains end-to-end double hydrogen bonds along the long axis of the molecule and side-to-side double hydrogen bonds along its short axis. These two-dimensional structures stack together to form the three-dimensional crystal. Takamizawa and Miyamoto found that in the β phase the terephthalamide molecules are more densely packed than in the α phase, but that the hydrogen-bond network is maintained. And this latter feature turns out to be the key to superelasticity.

The intermolecular forces that hold organic crystals together are usually much weaker than the interatomic covalent forces that bind together alloys and ceramics. However, in the present system, the collective hydrogen-bond network along the long and short axes of terephthalamide strengthens the otherwise weak intermolecular forces enough to prevent the crystal from fracturing on application of stress. More generally, the authors' study demonstrates the importance of hydrogen bonds in the supramolecular architectures of soft materials⁵. Because hydrogen bonds are much weaker than covalent bonds, supramolecular structures based on hydrogen bonds are more flexible against applied perturbations such as mechanical force, heat and light. This flexibility means that dissociation and association of the components that make up the supramolecular structure take place easily on application of such external stimuli, dissipating the applied perturbation smoothly⁶.

Soft superelastic materials could find several applications. For example, in microfluidic devices, the pressure of the fluid that flows in microchannels needs to be maintained below a crucial level to avoid damage to the channels. Generally, external pumps or internal valves control such pressure, but independent sensors are used to measure it.

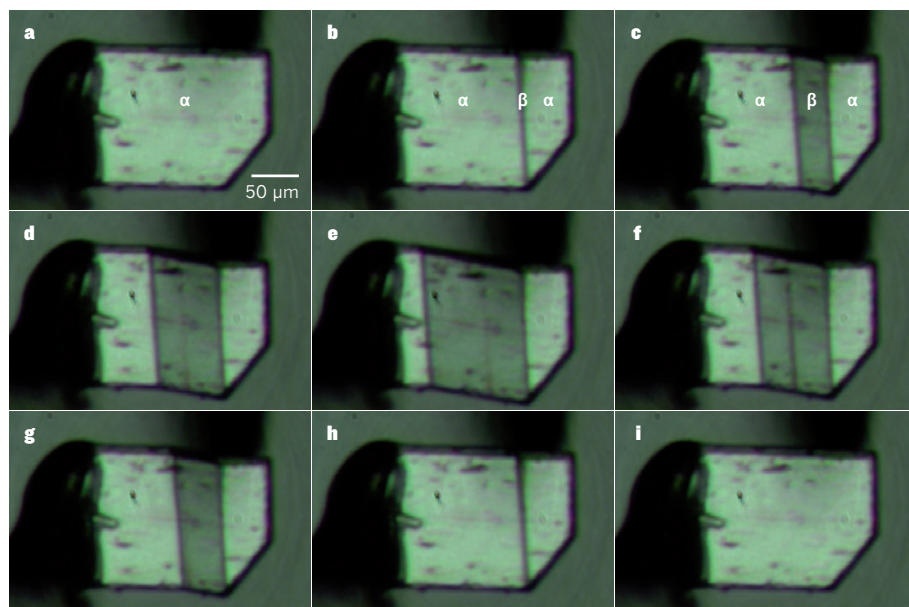


Figure 1 | Reversible deformation of a single organic crystal. Takamizawa and Miyamoto² pushed a metal blade against a single crystal of terephthalamide and observed how the crystal underwent reversible deformation. The crystal is initially in a crystallographic phase known as the α phase (a). When the blade is pushed against the crystal and the stress applied by it reaches a constant value, the crystal undergoes a phase transformation into a β phase at the contacting area between the blade and the crystal surface. This phase grows first along the pushing direction of the blade (b) and then perpendicularly to this direction, bending the crystal at the interface between the two phases (c–e). When the blade is pulled back, the crystal undergoes the reverse transformation (f–h), ultimately returning to its initial form (i). The top-right black region in these microscopy images is the blade, and the black region on the left is the glue used to fix the crystal to an underlying stand.

A superelastic organic crystal such as that presented here could be used to make internal valves that both sense and control the pressure in these devices. Such superelastic materials could also act as fillers in shock absorbers designed to dampen shock and vibration. ■

Tomiki Ikeda and Toru Ube are in the Research and Development Initiative, Chuo University, 1-13-27 Kasuga, Bunkyo-ku,

Tokyo 112-8551, Japan.
e-mail: tikeda@tamacc.chuo-u.ac.jp

1. Ölander, A. *J. Am. Chem. Soc.* **54**, 3819–3833 (1932).
2. Takamizawa, S. & Miyamoto, Y. *Angew. Chem. Int. Edn* **53**, 6970–6973 (2014).
3. Otsuka, K. & Ren, X. *Prog. Mater. Sci.* **50**, 511–678 (2005).
4. Jani, J. M., Leary, M., Subic, A. & Gibson, M. A. *Mater. Design* **56**, 1078–1113 (2014).
5. Kato, T., Mizoshita, N. & Kishimoto, K. *Angew. Chem. Int. Edn* **45**, 38–68 (2006).
6. Prins, L. J., Reinhoudt, D. N. & Timmerman, P. *Angew. Chem. Int. Edn* **40**, 2382–2426 (2001).

STEM CELLS

Reprogramming finds its niche

Production of blood stem cells from reprogrammed adult cells is notoriously difficult. It emerges that a supportive microenvironment may be crucial for their efficient generation. [SEE ARTICLE P.312](#)

DANIEL LUCAS & PAUL S. FRENETTE

One-marrow transplants can be life-saving, but a large proportion of patients who are in need of a transplant — particularly those from ethnic minorities — lack suitable donors. Blood-cell precursors called haematopoietic stem cells are the basis of transplants because, when they are injected intravenously, they can migrate and engraft into the bone marrow, regenerating every blood-cell lineage. One way to combat the donor deficit, therefore, would be to generate patient-derived haematopoietic stem cells. However, this strategy has been hampered by problems with engrafting engineered stem cells, and by difficulties with maintaining haematopoietic ‘stemness’ in laboratory-cultured cells. On page 312 of this issue, Sandler *et al.*¹ describe an approach for generating haematopoietic stem cells that circumvents these problems.

In their seminal experiment², the stem-cell biologists Shinya Yamanaka and Kazutoshi Takahashi reprogrammed skin fibroblast cells into a ‘reset’ state. Starting with a series of candidate transcription factors, the researchers defined a combination of four factors that induce complete cellular dedifferentiation. The reprogrammed cells, called induced pluripotent stem (iPS) cells, can theoretically differentiate into any cell type in the body. However, differentiation of iPS cells into functional adult tissues has proved to be a challenge, owing to our lack of understanding about the complex cues required to program cells *in vitro*. As such, differentiation protocols for haematopoietic stem cells (HSCs) tend to yield embryonic-like blood cells that do not

engraft efficiently into bone marrow³.

An alternative strategy is the direct reprogramming of adult cells into another lineage, without going through a pluripotent-cell stage. Adult fibroblasts have been successfully reprogrammed into several cell types, including neurons, cardiomyocytes and hepatocytes⁴. Last year⁵, four transcription factors (Gata2, cFos, Gfi1b and Etv6) were used to reprogram mouse fibroblasts into cells that expressed HSC surface markers and differentiated into blood-cell progenitors *in vitro* (Fig. 1). However, the reprogrammed cells could not robustly engraft into bone marrow after transplantation.

During embryonic development, HSCs arise from vascular cells that line the aorta, and the cells continue to require signals from the vascular bed, or niche, for their maintenance and function throughout their lives. Sandler *et al.* reasoned that they could enhance the efficiency of direct reprogramming and maintain the self-renewing abilities of the induced HSCs (iHSCs) by starting with a cell type with a similar developmental origin to HSCs, and growing the cells in a microenvironment comparable to their *in vivo* niche.

The authors isolated human umbilical-vein endothelial cells (HUVECs, readily available cells that line the umbilical vein), and forced them to express 26 transcription factors that are enriched in HSCs, but not in HUVECs. The researchers maintained the cultured cells in a medium that lacked serum, which can impair HSC maintenance (serum is normally included in culture media because it contains growth factors that promote cell proliferation). Sandler and colleagues kept the cells on a feeder-cell layer; this underlying



50 Years Ago

Some top-rank public schools and university colleges produce men of brilliant academic achievement who have poor judgement, no power of decision and no capacity to delegate work or to control men. These men can be the tragedies of industry because their deficiencies are not revealed in their academic record and are difficult to detect at a selection interview. They can get started on a promising career, but end in the wilderness of the unpromotable clever boys ... some of the highest places in industry have been filled successfully by men whose education has been obtained the hard way. In these cases, the task of getting education and training the hard way has imposed personal disciplines which have probably led imperceptibly to the acquisition of those characteristics needed in industry. Sometimes, however, such a course produces an almost characterless ‘swot’.

From *Nature* 18 July 1964

100 Years Ago

Everyone is familiar with the dramatic story of Bernard Palissy, the potter, and how he fired a kiln with his household furniture in order to produce sufficient heat to melt his glazes, but his scientific work is rarely mentioned ... during the years 1575–84 he exercised great influence upon society in the city. He lectured in agriculture, chemistry, mineralogy, and geology, and illustrated his lectures with demonstrations of natural objects from his museum. “Into the faces of the learned of his time he thrust his facts; he urged the might of the verified fact, the tests of practical experience, the demonstration of the senses; and these in a keen and original way.” ... At the age of eighty Palissy was thrown into the Bastille as a dangerous heretic.

From *Nature* 16 July 1914

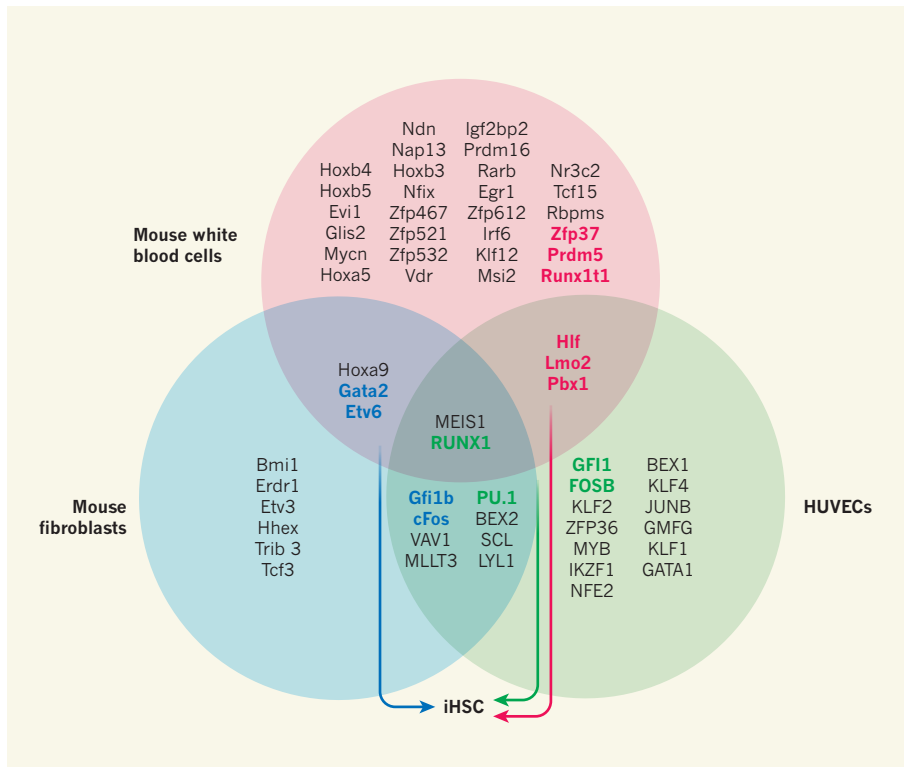


Figure 1 | In search of factors that induce haematopoietic stem cells. A Venn diagram illustrates the overlapping groups of transcription factors that have been tested in the quest to generate induced haematopoietic stem cells (iHSCs) from three adult cell types — mouse fibroblasts⁵ (the transcription-factor pool tested is represented by the blue circle), mouse white blood cells⁷ (the pink circle) and human umbilical vein endothelial cells (HUVECs, green)¹. Stepwise elimination of the transcription factors that were unnecessary for each protocol led to the identification of distinct groups that could generate iHSCs (bold text, colour indicates which transcription-factor combinations successfully reprogrammed each cell type). Sandler *et al.*¹ report that the production of iHSCs from HUVECs requires four transcription factors: FOSB, GFI1, RUNX1 and PU.1.

cell monolayer released factors that made the culture conditions similar to the micro-environment of the HSC niche. The feeder cells, called E4ECs, were endothelial cells engineered to overexpress an adenoviral gene, *E4ORF1*, that promotes their survival, but not their proliferation, thereby maintaining a state that mimics the niche⁶.

When HUVECs were cultured in these conditions, Sandler and colleagues found that a small subset could form haematopoietic colonies. Systematic elimination of transcription factors that were unnecessary for reprogramming revealed that a combination of 4 of the 26 factors — FOSB, GFI1, RUNX1 and PU.1 — could reprogram HUVECs (Fig. 1). To be successful, reprogramming must simultaneously suppress the original cellular identity and confer a new one. The authors speculate that PU.1 combined with GFI1 downregulated vascular genes, possibly in combination with FOSB, and that PU.1 and RUNX1 upregulated haematopoietic-specifying genes. Reprogrammed HUVECs became self-renewing HSCs that could serially engraft into the bone marrow of immunodeficient mice and differentiate into mature blood cells.

Earlier this year⁷, another group reported

the transformation of mature white blood cells from mice into engraftable HSCs that can form all blood-cell lineages. Reprogramming was accomplished with six transcription factors (Runx1t1, Hlf, Lmo2, Prdm5, Pbx1 and Zfp37), and the cells were matured *in vivo* to generate iHSCs (Fig. 1). It is probable, although it has not yet been demonstrated, that maturation in native HSC niches promoted cell survival and provided cues for iHSC generation.

Surprisingly, comparison of Sandler and colleagues' protocol with those used to reprogram mouse white blood cells⁷ or fibroblasts⁵ reveals that each method used a different transcription-factor cocktail to generate iHSCs. This may result from species differences, from the ability of each cell type to respond to different transcription factors or from the different epigenetic state of each cell type — that is, genomic modifications that affect gene expression without changing DNA sequence. Consistent with this possible role for epigenetic state, Sandler and colleagues' transcription-factor cocktail could not reprogram endothelial cells derived from embryonic stem cells, but could reprogram adult dermal microvascular endothelial cells.

It is worth mentioning that the surprisingly limited overlap in reprogramming factors between the three studies^{1,5,7} is representative of the differing starting pools of transcription factors used. Indeed, even though each was chosen on the basis of selective expression in HSCs, only two factors (RUNX1 and MEIS1) were included in the initial pool of every study. Thus, definitive conclusions about the cell-specificity and transcription-factor requirements for generating iHSCs await further analyses. The fact that the same result can be achieved with three different molecular combinations suggests a multiplicity of options for generating iHSCs.

The ability to reprogram adult endothelial cells has exciting implications for gene editing and cell therapy for blood diseases. Although HSCs have always been a desirable target for gene therapy, the difficulties of maintaining them in culture have limited their use. As adult endothelial cells can be cultured for several days without apparent loss of reprogramming efficiency, one can predict that patient-specific endothelial cells could be purified, genetically corrected, selected and then reprogrammed to deliver functional iHSCs.

As with all stem cells reprogrammed in culture, the risk of cancerous transformation remains. Although Sandler and colleagues found no signs of transformation 10 months after transplanting the iHSCs into mice, most of the factors used in iHSC generation are also associated with the development of leukaemia. This highlights the thin line between promoting self-renewal of healthy HSCs and potentiating cancerous transformation. A greater understanding of the reprogramming mechanisms at play may overcome this potential problem. Furthermore, such understanding will produce much-needed insight into the signals that trigger HSC emergence, and the molecular networks that instruct HSC programming. ■

Daniel Lucas is in the Department of Cell and Developmental Biology, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA.

Paul S. Frenette is at the Ruth L. and David S. Gottesman Institute for Stem Cell and Regenerative Medicine Research, Albert Einstein College of Medicine, New York, New York 10461, USA.

e-mail: paul.frenette@einstein.yu.edu

1. Sandler, V. M. *et al.* *Nature* **511**, 312–318 (2014).
2. Takahashi, K. & Yamanaka, S. *Cell* **126**, 663–676 (2006).
3. Sturgeon, C. M., Ditadi, A., Clarke, R. L. & Keller, G. *Nature Biotechnol.* **31**, 416–418 (2013).
4. Pfaff, N. & Cantz, T. *Cell Stem Cell* **13**, 131–133 (2013).
5. Pereira, C.-F. *et al.* *Cell Stem Cell* **13**, 205–218 (2013).
6. Butler, J. M. *et al.* *Blood* **120**, 1344–1347 (2012).
7. Riddell, J. *et al.* *Cell* **157**, 549–564 (2014).

This article was published online on 2 July 2014.

A deep crust–mantle boundary in the asteroid 4 Vesta

Harold Clenet¹, Martin Jutzi², Jean-Alix Barrat³, Erik I. Asphaug⁴, Willy Benz² & Philippe Gillet¹

The asteroid 4 Vesta was recently found to have two large impact craters near its south pole, exposing subsurface material. Modelling suggested that surface material in the northern hemisphere of Vesta came from a depth of about 20 kilometres, whereas the exposed southern material comes from a depth of 60 to 100 kilometres. Large amounts of olivine from the mantle were not seen, suggesting that the outer 100 kilometres or so is mainly igneous crust. Here we analyse the data on Vesta and conclude that the crust–mantle boundary (or Moho) is deeper than 80 kilometres.

Global mapping with high-resolution imagery by the Dawn probe revealed that the south polar depression is composed of two overlapping impact basins, Veneneia and Rheasilvia¹. This discovery is critical in the search for the Mohorovičić discontinuity (Moho). Indeed, although a single impact is expected to excavate rocks only from the crust, recent numerical simulations^{2,3} taking into account both sequential events show that excavation and ejection of mantle material during the second impact would be facilitated because the first one would have already thinned or removed the crust locally.

Impact simulations in three dimensions have been able to reproduce Vesta's topography accurately². The results of this model allow the source depths or provenances of rocks to be directly investigated today. Two distinct sets of observables are considered for comparison with modelling observations: the surface of Vesta, which includes the material outcropping in the basins and the ejecta covering the rest of the asteroid, and the meteoroids and asteroids that escaped during the impacts and are the probable source of the howardite–eucrite–diogenite (HED) meteorites originating from Vesta⁴.

Mapping the predicted provenance of surface material (Fig. 1) shows that a large amount of the rocks exposed in the south pole region should come from depths exceeding 50 km. Simulations predict initial depths of up to ~60–100 km in the central mound of Rheasilvia and in the region where the impact basins overlap². If the crust of Vesta is ~30–40 km thick, as proposed in magma-ocean crystallization models^{5–8}, a succession of two impacts would have dug well into the mantle, producing large outcrops of olivine-rich rocks within the basins.

Mineralogical mapping of Vesta's surface with images from the VIR instrument onboard the Dawn probe shows that pyroxenes are ubiquitous in the southern hemisphere, while no olivine is observed^{9–11}, even where rocks come from the deepest levels in numerical simulations (Fig. 1). Admittedly, the mantle spectral signature could have been partially masked by late-impact gardening. But because the mantle is highly enriched in olivine, and because the outcrops occur over a broad expanse, some pixels should exhibit a definitive olivine signature. The conclusion that olivine does not represent a large mineral fraction of the rocks¹² is at odds with the higher content expected in deep mantle rocks, and argues against the idea that mantle is excavated and exposed by the successive impacts.

The HED meteorites¹³ are a large collection of basaltic and ultramafic samples that originated from Vesta. There is no reason, a priori, why their relative proportions should equal the proportions of Vesta's surface covered by the different lithologies¹³. They come from the small asteroids

(vestoids) that were ejected by the large impacts¹⁴, so their proportions will, if anything, be representative of the lithologies that escaped Vesta. The amount of material escaping Vesta following both collisions can be reliably estimated² (see Methods). The first impact ejected away material from up to about 25 km deep (Fig. 2). The second overlapping impact dug deeper, up to 20–80 km below the original surface, because Rheasilvia formed on top of rocks excavated by Veneneia. A third of all the escaping material comes from depths greater than 40 km. So, if the assumption of a thin crust^{7,8} is considered, samples of mantle should be found in the HED suite.

While the details of the mass distribution shown in Fig. 2 depend on the impact geometry, the overall result is robust because it is closely tied to the impact suite (model results) that best explains the detailed shape of Vesta. In addition, two distinct small asteroid populations are predicted, one for each of the massive cratering events, which is consistent with recent evidence for spectral colour variations within the Vesta family¹⁵. However, it is not clear to what degree the original proportions of the escaping materials have been changed during the dynamical and collisional evolution of the fragments. The total amount of material escaping Vesta as a result of the two giant impacts is approximately 2.7×10^{18} kg, according to smooth particle hydrodynamics (SPH) modelling², which is consistent with geological estimates of basin formation¹. This exceeds the observational estimates of the total mass of the vestoids ($0.5\text{--}3 \times 10^{17}$ kg³). While these estimates do depend on the assumed size distribution³, this does not change the fact that far more material escaped Vesta than is observed today.

This implies that the vestoids have been greatly eroded, as was already predicted by detailed simulations of the dynamical evolution of the Vesta family¹⁶. But there is no reason why this depletion would deplete the olivines—in escaping fragments of the deep mantle—and not the pyroxenes. If Vesta's mantle was originally about 40 km deep, then a third of the material escaping Vesta should have olivine-rich lithologies (see Fig. 2), and a significant fraction of the HEDs should come from deeper still. Olivine should be reflected in the composition of the main-belt and near-Earth vestoids as well as in the HED meteorite suite.

Only a few meteorites enriched in olivine have been collected, and they do not originate from the mantle, but instead formed in plutons¹⁷. One explanation for the 'missing' olivine is that only ejecta from the first, shallower-digging impact reached the Earth. If so, then olivine-rich deeper rocks excavated during the second impact should still be prominently visible among the vestoids. This is definitely not the case, as all of them have

¹EPFL, Institute of Condensed Matter Physics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Station 3, CH-1015 Lausanne, Switzerland. ²Physics Institute, Space Research and Planetary Sciences, Center for Space and Habitability, University of Bern, Sidlerstrasse 5, 3012 Bern, Switzerland. ³Université de Bretagne Occidentale, Institut Universitaire Européen de la Mer, CNRS UMR 6538, Place Nicolas Copernic, 29280 Plouzané, France. ⁴School of Earth and Space Exploration, Arizona State University, PO Box 876004, Tempe, Arizona 85287, USA.

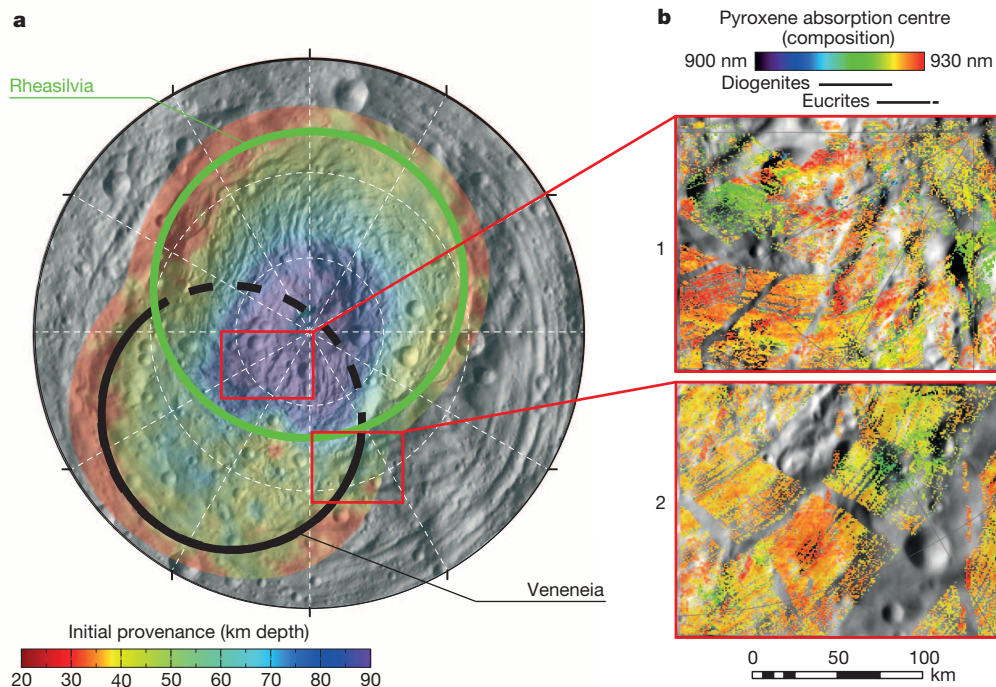
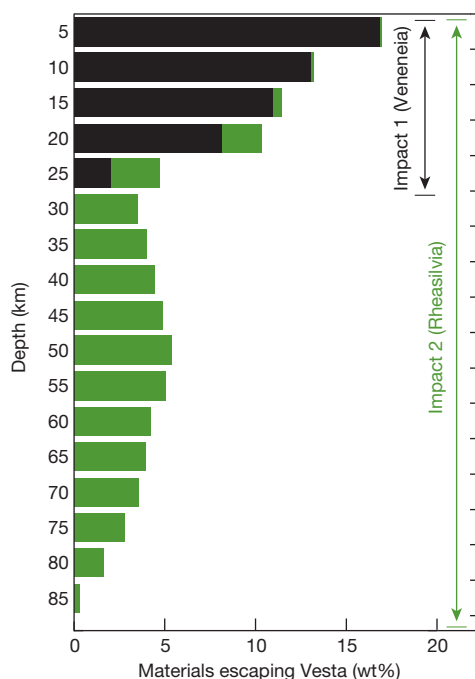


Figure 1 | Pyroxenes composition in regions expected to expose mantle rocks. **a**, Initial provenance (depth) of the exposed material on the surface (from numerical simulations²). The rims of the impact basins Rheasilvia and Veneneia are outlined. **b**, Absorption centres from modified Gaussian modelling²⁸, showing the composition of the pyroxenes. The colours in the two maps in **b** give the absorption centres calculated from MGM for each pixel (results fall between 900 nm and 930 nm). The relation between absorption centres and composition is given by the diogenite and eucrite ranges below the coloured scale (following ref. 9). Eucrites have absorption centres shifted more

spectral signatures similar to eucrites, diogenites or a mixture of both^{18–20}. Once again, those observations are at odds with the idea of an approximately 30–40-km-thick crust.



towards 930 nm than diogenites do and thus cyan and green pixels correspond to diogenites while yellow/orange/red pixels correspond to eucrites. Region 1 (upper red box in **a**), within the two basins and encompassing a portion of Rheasilvia's central peak, is where rocks come from the deepest levels in simulations. Region 2 (lower red box in **a**) shows a diogenite-like lithology⁹, suggested to be a major constituent of the uppermost mantle¹¹. As in previously published results^{9,25}, an olivine content higher than the detection threshold is detected nowhere, indicating that no mantle rocks are outcropping.

The lack of olivine detections in the Veneneia/Rheasilvia region, and the simultaneous lack of mantle samples among the vestoids/HEDs, together provide evidence that the Moho was not reached during the two impacts. Consequently, the crust of Vesta must be much thicker than 40 km, and possibly as thick as 80 km, according to the cratering simulations. While this is in agreement with the interpretation of gravity data²¹, magma-ocean crystallization models are unable to reproduce such a thick crust^{7,8}. Indeed, models are required to explain both the oxygen isotopic homogeneity of Vesta and the trace-element features of diogenites²².

The homogeneity of oxygen isotopic signatures in HEDs attests to a global-scale melting event following accretion²². It has been suggested that the deep cumulates formed during the cooling of the magma ocean suffered extensive remelting. The resulting melts could have formed diogenitic intrusions within the massive eucritic crust^{23,24} (Fig. 3). The probability of the existence of such crustal intrusions is strengthened by the recent discovery of scattered patches of olivine-rich rocks (50%–80% olivine), hundreds of metres in size, that occur over an area of about a hundred kilometres square in the northern hemisphere of Vesta²⁵. As those outcrops are found around impact craters too small to reach the mantle, they might indicate the exhumation of upper crustal plutons, with locally olivine-enriched layers, rather than exposures of a global olivine-rich layer.

Figure 2 | Initial depths and mass fractions of rocks that escaped Vesta. Escaped rocks provides a unique sampling profile of the interior of Vesta. They should be statistically represented in the HED meteorites suite. The relative proportion of material that escaped Vesta, compared to the total mass loss, is given as a function of its original depth before the impacts. Mass fractions and depths are obtained using previously published three-dimensional numerical simulation of impacts². The first impact (Veneneia, in black) ejected mostly material from shallow depths (<25 km) while the second one (Rheasilvia, in green) ejected material from greater depths (mainly between 20 km and 80 km).

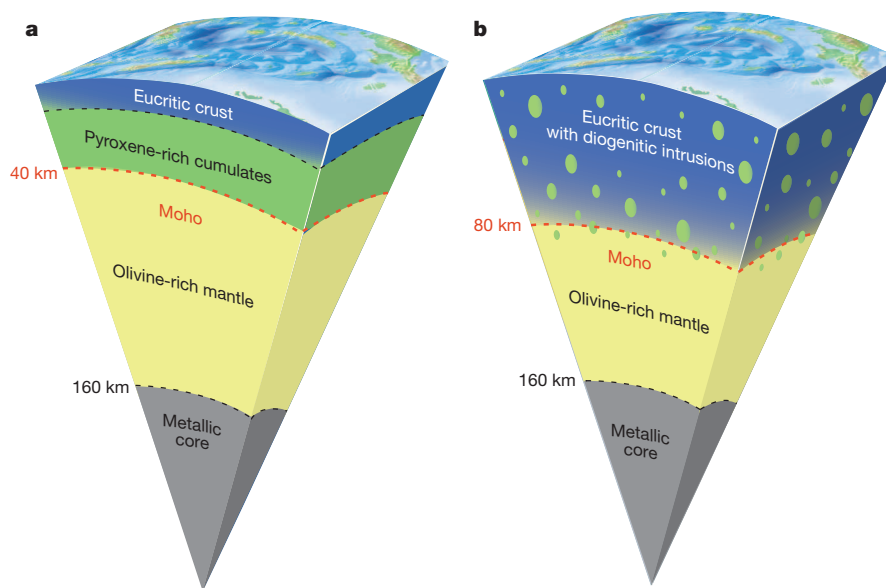


Figure 3 | Potential internal structures for Vesta. **a**, Classical model of the internal layering of Vesta resulting from the crystallization of a magma ocean, generally assuming the crust to be around 40 km thick⁷. **b**, The scenario of eucritic crust intruded by plutons, which leads to a much thicker crust²⁴. Diogenitic plutons can have locally olivine-enriched layers. The depth of the core is 160 km (ref. 29).

It has been known since the 1970s that eucrites are poor in sodium and other volatile elements²⁶. This might imply that Vesta formed from volatile-poor dusts in an incompletely condensed solar nebula²⁷, or via a complex path related to inefficient accretion. With a deep Moho, it appears clear that the mantle is much thinner than expected (Fig. 3), leading to the conclusion that Vesta contains far less olivine than predicted by chondritic models⁸. This could be additional evidence that its bulk chemical composition deviates substantially from a chondritic composition for major elements as well.

METHODS SUMMARY

We here analyse data presented in previous papers^{2,9}. We use the results of recent three-dimensional SPH modelling² to track the dynamical evolution and redistribution of the material during the simulations. Using the initial location of each particle before the impacts, we define the initial depth as the radial distance to the surface and we calculate the provenance of the material that is finally excavated onto the surface or lost from the asteroid (that is, that reach a speed higher than the escaped velocity).

We also analyse data acquired in the southern hemisphere of Vesta by the Dawn VIR instrument. All available images were first processed following previously published pipelines^{9,10}, which include ISIS3 (<http://isis.astrogeology.usgs.gov/>) and in-house procedures to correct for photometry, bad pixels and spatial misalignment. We thus produce a global mosaic with a resolution of ten pixels per degree. We then used a recent approach based on the Modified Gaussian Model (MGM)²⁸ to detect low olivine content in mixtures with pyroxene.

Before any systematic mapping, we additionally tested the capability of the chosen approach to detect large amounts of olivine in HEDs. On the two olivine-diogenites tested (with contents of, respectively, 50% and 57% of olivine), an olivine/pyroxene mixture was successfully detected. A certified limit of about 50% is enough to definitely spot any mantle outcrop (60%–80% olivine) on Vesta's surface, and yet our analyses of VIR data reproduce previous findings where pyroxenes are ubiquitous in the southern hemisphere, but no olivine is observed^{9–11}.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 February; accepted 16 April 2014.

1. Schenk, P. *et al.* The geologically recent giant impact basins at Vesta's south pole. *Science* **336**, 694–697 (2012).
2. Jutzi, M., Asphaug, E., Gillet, P., Barrat, J.-A. & Benz, W. The structure of the asteroid 4 Vesta as revealed by models of planet-scale collisions. *Nature* **494**, 207–210 (2013).
3. Ivanov, B. A. & Melosh, H. J. 2D numerical modeling of the Rheasilvia impact formation. *J. Geophys. Res. Planets* **118**, 1545–1557 (2013).

4. McCord, T. B., Adams, J. B. & Johnson, T. V. Asteroid Vesta: spectral reflectivity and compositional implications. *Science* **168**, 1445–1447 (1970).
5. Ruzicka, A., Snyder, G. A. & Taylor, L. A. Vesta as the howardite, eucrite and diogenite parent body: implications for the size of a core and for large-scale differentiation. *Meteorit. Planet. Sci.* **32**, 825–840 (1997).
6. Righter, K. & Drake, M. J. A magma ocean on Vesta: core formation and petrogenesis of eucrites and diogenites. *Meteorit. Planet. Sci.* **32**, 929–944 (1997).
7. Mandler, B. E. & Elkins-Tanton, L. T. The origin of eucrites, diogenites, and olivine diogenites: magma ocean crystallization and shallow magma chamber processes on Vesta. *Meteorit. Planet. Sci.* **48**, 1–17 <http://dx.doi.org/10.1111/maps.12135> (2013).
8. Toplis, M. J. *et al.* Chondritic models of 4 Vesta: implications for geochemical and geophysical properties. *Meteorit. Planet. Sci.* **16**, 1–16 (2013).
9. Ammannito, E. *et al.* Vestan lithologies mapped by the visual and infrared spectrometer on Dawn. *Meteorit. Planet. Sci.* **48**, 1–14 <http://dx.doi.org/10.1111/maps.12192> (2013).
10. De Sanctis, M. C. *et al.* Spectroscopic characterization of mineralogy and its diversity across Vesta. *Science* **336**, 697–700 (2012).
11. McSween, H. Y. *et al.* Composition of the Rheasilvia basin, a window into Vesta's interior. *J. Geophys. Res.* **118**, 1–12 (2013).
12. Beck, A. W. *et al.* Challenges in detecting olivine on the surface of 4 Vesta. *Meteorit. Planet. Sci.* **48**, 1–11 <http://dx.doi.org/10.1111/maps.12160> (2013).
13. McSween, H. Y. *et al.* Dawn; the Vesta-HED connection; and the geologic context for eucrites, diogenites, and howardites. *Meteorit. Planet. Sci.* **48**, 2090–2104 (2013).
14. Binzel, R. P. & Xu, S. Chips off of asteroid 4 Vesta: evidence for the parent body of basaltic achondrite meteorites. *Science* **260**, 186–191 (1993).
15. Bus, S. J. Evidence for spectral color variation within the Vesta family. In *8th Workshop on 'Catastrophic Disruption in the Solar System'* http://www.cd8.hawaii-conference.com/wp-content/uploads/2013/06/CD8_abs_Bus.pdf (2013).
16. Nesvorný, D. *et al.* Fugitives from the Vesta family. *Icarus* **193**, 85–95 (2008).
17. Beck, A. W. & McSween, H. Y. Diogenites as polymict breccias composed of orthopyroxenite and harzburgite. *Meteorit. Planet. Sci.* **45**, 850–872 (2010).
18. Mayne, R. G., Sunshine, J. M., McSween Jr, H. Y., Bus, S. J. & McCoy, T. J. The origin of Vesta's crust: insights from spectroscopy of the Vestoids. *Icarus* **214**, 147–160 (2011).
19. Reddy, V., Nathues, A. & Gaffey, M. J. First fragment of asteroid 4 Vesta's mantle detected. *Icarus* **212**, 175–179 (2011).
20. Buratti, B. J. *et al.* Vesta, vestoids, and the HED meteorites: interconnections and differences based on Dawn Framing Camera observations. *J. Geophys. Res.* **118**, 1991–2003 (2013).
21. Park, R. S. *et al.* Gravity field expansion in ellipsoidal harmonic and polyhedral internal representations applied to Vesta. *Icarus* <http://dx.doi.org/10.1016/j.icarus.2013.12.005> (in the press).
22. Greenwood, R. C. *et al.* The oxygen isotope composition of diogenites: evidence for early global melting on a single, compositionally diverse, HED parent body. *Earth Planet. Sci. Lett.* **390**, 165–174 (2014).
23. Barrat, J.-A., Yamaguchi, A., Zanda, B., Bollinger, C. & Bohn, M. Relative chronology of crust formation on asteroid Vesta: insights from the geochemistry of diogenites. *Geochim. Cosmochim. Acta* **74**, 6218–6231 (2010).
24. Yamaguchi, A., Barrat, J.-A., Ito, M. & Bohn, M. Post-eucritic magmatism on Vesta: evidence from the petrology and thermal history of diogenites. *J. Geophys. Res.* **116**, E08009 (2011).

25. Ammannito, E. *et al.* Olivine in an unexpected location on Vesta's surface. *Nature* **504**, 122–125 (2013).
26. Tera, F., Eugster, O., Burnett, D. S. & Wasserburg, G. J. Comparative study of Li, Na, K, Rb, Cs, Ca, Sr and Ba abundances in achondrites and in Apollo 11 lunar samples. *Proc. Apollo 11 Lunar Sci. Conf.* **2**, 1637–1657 (1970).
27. Hans, U., Kleine, T. & Bourdon, B. Rb–Sr chronology of volatile depletion in differentiated protoplanets: BABI, ADOR and ALL revisited. *Earth Planet. Sci. Lett.* **374**, 204–214 (2013).
28. Clénet, H. *et al.* A new systematic approach using the Modified Gaussian Model: insight for the characterization of chemical composition of olivines, pyroxenes and olivine–pyroxene mixtures. *Icarus* **213**, 404–422 (2011).
29. Russell, C. T. *et al.* Dawn at Vesta: testing the protoplanetary paradigm. *Science* **336**, 684–686 (2012).

Acknowledgements M.J. acknowledges support from the Swiss National Science Foundation through the Ambizione program. J.-A.B. acknowledges support from the INSU Programme National de Planétologie. E.I.A. was sponsored by the NASA Planetary Geology and Geophysics Program.

Author Contributions H.C. analysed data and led the research. M.J. performed the numerical simulations. P.G. initiated the collaboration and funded part of the research. All authors interpreted the results and contributed to the preparation of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.C. (harold.clenet@epfl.ch).

METHODS

Numerical simulations of impact. We here use the results of previously published three-dimensional SPH simulations². In the three-dimensional modelling of the two subsequent impacts, the dynamical evolution and redistribution of the material (SPH particles) was tracked during the simulation. Note that self-gravity was computed throughout the whole simulation. The provenance of surface material shown in Fig. 1 (left) is computed in the same way as in ref. 2. However, here we use a stereographic projection, which corresponds to the same projection as the underlying background.

Using the initial location (before the impacts) of each SPH particle we define the initial depth as the radial distance to the surface of the initially spherical target. To determine whether or not a particle will escape Vesta owing to an impact, we compare its ejection velocity v_{esc} with the escape speed of Vesta, using $v_{\text{esc}} = 360 \text{ m s}^{-1}$. The total mass of ejecta originating from a certain layer of the target is then given by the summation of all particles located within this layer, which have ejection velocities $v_{\text{eject}} > v_{\text{esc}}$. This procedure is used to analyse both the Veneneia and the Rheasilvia impacts. It can be noted that an increase of the required escape speed leads to a global decrease of the total amount of material escaping Vesta, without, however, affecting the relative distribution among the initial depths.

Although the simulations reproduce well the observed topography of Vesta², different initial conditions might lead to a good match as well. The uncertainties are due to the pre-impact shape, the impact angle and velocity of each impact and Vesta's rotation axis, which are all unknown. However, the overall results of our analysis (the provenance of ejecta from deep layers) are expected to be robust since they are mainly produced by the significant overlap of the two giant basins. Moreover, the SPH modelling results are roughly consistent with the findings of ref. 3 when the subsequent formation of both basins is considered.

Processing of Dawn VIR data. We process all the available Dawn VIR³⁰ images over the southern hemisphere from the High Altitude Mapping Orbit (HAMO) 1 and 2 and from the Low Altitude Mapping Orbit (LAMO) to produce a mosaic similar to the ones published in refs 9 and 10. Images were downloaded from NASA's Planetary Data System Small Bodies Node (<http://pds-smallbodies.astro.umd.edu/>) with level 1B calibration. They were first processed through the classical ISIS3 pipeline³¹ (details for each function can be found in the online ISIS3 manual available at <http://isis.astrogeology.usgs.gov/Application/index.html>). Each data cube is read with the *dawnvir2isis* procedure. Ground positions and photometric viewing angles are computed using the *spiceinit* function. In parallel, the associated quality and geometry cubes are produced using the *pds2isis* and *phocube* functions respectively. Then pixels where observation angles are too high ($>75^\circ$) are removed using *photrim*. Finally, we apply a photometric correction with the *photomet* function (in HapkeHen mode) and using the parameters found in the literature^{32–34} (macroscopic roughness parameter $\theta = 20$, single-scattering albedo $W_h = 0.52$, single-term Henyey–Greenstein coefficient $Hg1 = -0.29$, width of the opposition surge $Hh = 0.04$, amplitude of the opposition surge $B0 = 1.03$).

We then use in-house routines for additional processing. We first filter the bad pixels in each image using the associated quality cube. Because the visible and near-infrared parts of an image are acquired with two distinct detectors³⁰, geographical misalignments could exist and must be taken into account before the fusion of the two spectral domains. To correct this geographical misalignment, we use the method of the Dawn team⁹. We compute for each detector the latitude/longitude coordinates of the pixels in the corners of the image and we apply, if needed, a simple translation on the visible part to match the infrared coordinates. Images with non-homogenous misalignment are removed from the final mosaic. The conversion from radiance to the irradiance/solar flux is done using Kurucz solar irradiance spectrum resampled at VIR-infrared sampling and resolution. Finally, all the VIR images are projected using the geometry information and assembled in a mosaic (resolution of ten pixels per degree) covering all of the southern hemisphere of Vesta.

The quantitative interpretation of mineralogy from spectra is limited as it is hampered by the overlap of the absorption features, particularly when there is a mixture of two or three minerals^{35–37}. The Modified Gaussian Model (MGM)^{38–40} aims at deconvolving overlapping absorptions of mafic mineral spectra into their fundamental absorption components. It is achieved by considering a sum of modified Gaussian functions characterized by their band centres, widths and strengths. The specific aim of this model is to account directly for electronic transition processes³⁸.

However, MGM results are sensitive to the initial parameters²⁸ and thus it cannot be implemented blindly on an entire data set as acquired on Vesta's surface. An automatic procedure has been implemented to deal with unknown mafic mineralogy in the case of natural rock spectra²⁸. An automatic analysis of the shape of the spectrum is first performed (spectrum maxima and minima are used to estimate, to first order, the absorption strengths and widths). The continuum is handled with a second-order polynomial initially adjusted on the local maxima along the spectrum (curvature, slope and shift are free to move during the modelling). All the mixture

possibilities involving orthopyroxene, clinopyroxene and olivine are considered²⁸ and, accordingly, different numbers of Gaussians (from 3 to 7), depending on the potential complexity of the mixture, are used for each of the seven configurations. Additional Gaussians centred around 0.5 μm (ultraviolet charge-transfer absorption), 0.6 μm (ferric absorption) and at 1.4 μm , 1.9 μm and 2.3 μm (hydration and alteration effects) may be requested to account for spectral features not related to mafic mineralogy. The initial settings for the three parameters for each Gaussian for the seven different configurations are made each time on the basis of the spectrum shape and the laboratory results available in the literature in the case of simple mixtures of mafic minerals²⁸.

Considering all the mixture possibilities with the three mafic components, MGM modelling is run seven times, with seven different initializations, on a given pixel. Root-mean-square residuals cannot be used as the only parameter to check for the validity of the results because a large number of Gaussian functions may result in low root-mean-square mathematical solutions without any physical meaning. Consequently, the returned MGM solutions are then assessed on the basis of a mineralogical sorting (that is, each modelled Gaussian functions must verify the spectroscopic criteria defined in the literature^{35,39,40}) and are accordingly either validated or discarded. Finally, the solutions kept are interpreted in terms of mineralogy²⁸.

The uncertainties on the calculated band centres have been determined on laboratory spectra to be $\pm 8 \text{ nm}$ in the 1- μm domain (and $\pm 17 \text{ nm}$ in the 2- μm domain, which is not used here)⁴¹. Those uncertainties are relevant because the absorption depths in the VIR data are comparable to the depths observed in the laboratory³.

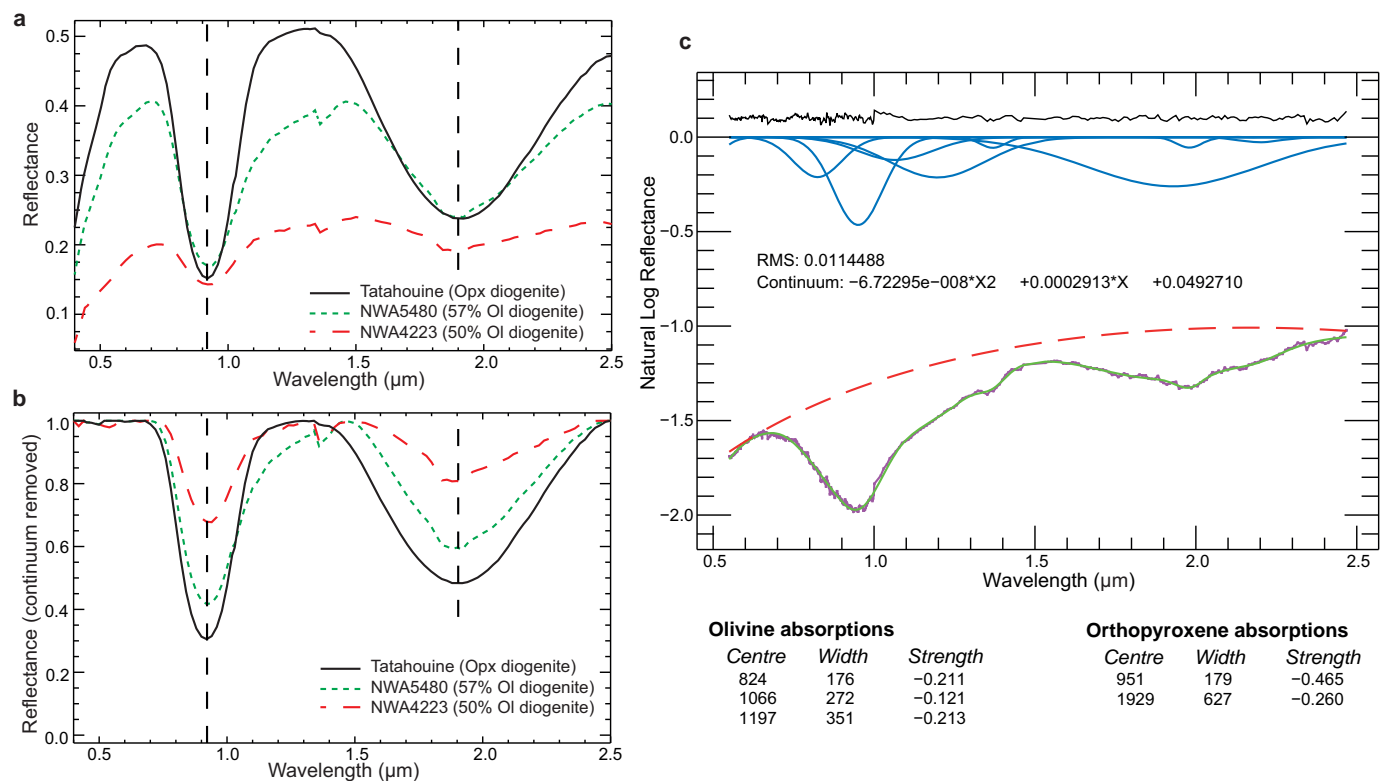
The adapted MGM approach is able to model both simple and complex mafic mineralogies, including binary and ternary mixtures (involving orthopyroxene, spectral type B clinopyroxene and olivine) for a large range of grain sizes. It has been extensively validated on a large range of laboratory spectra, in particular on some that are representative of HED compositions (various olivine-orthopyroxene mixtures with similar chemistry)²⁸. It was originally shown to be able to detect an olivine content as low as 10% to 50% on laboratory spectra²⁸, the exact limit depending mainly on the chemical composition and the grain size of each minerals in the mixture. The adapted MGM approach has also been validated in natural conditions on Earth⁴², the Moon⁴³ and Mars⁴⁴, each time allowing the detection of olivine with or without pyroxenes.

Nevertheless, before any systematic mapping on Vesta's surface, we choose to test the capability of this approach to detect significant amount of olivine on HEDs. We use two spectra from the meteorites NWA4223 and NWA5480 (ref. 45). Both are olivine-diogenites with olivine contents of 50% and 57%, respectively. We note that both spectra also show effects of terrestrial weathering, which can affect the spectral slope. Nevertheless, the continuum allows us to take into account this effect in the MGM approach and Earth validation has shown that band centre results remain reliable⁴².

The spectra clearly exhibit strong pyroxene signatures (Extended Data Fig. 1); however, in both cases MGM results confirm the detection of an olivine/pyroxene mixture, thus verifying the ability of the approach to detect systematically contents of at least 50% olivine. Magma-ocean crystallization models and simple mass-balance and thermodynamic constraints agree on the harzburgitic nature ($>40\%$ olivine mixed with orthopyroxene) of Vesta's mantle^{7,8}. The predicted olivine content lies within the range 60%–80%, which differs significantly from the low olivine abundances observed in the lithologies from the pyroxene-rich crust (olivine-diogenites generally contain less than 30% of olivine^{12,17}). Therefore, a certified limit of $\sim 50\%$ is enough to definitely spot any mantle outcrop (60%–80% olivine) on Vesta's surface.

30. De Sanctis, M. C. *et al.* The VIR spectrometer. *Space Sci. Rev.* **163**, 329–369 (2011).
31. Anderson, J. *et al.* Isis cartographic tools for the Dawn Framing Camera and Visual and Infrared Spectrometer. *AGU Fall Meet. Abstr.* U31A-0009 (American Geophysical Union, 2011).
32. Reddy, V. *et al.* Color and albedo heterogeneity of Vesta from Dawn. *Science* **336**, 700–704 (2012).
33. De Sanctis, M. C. *et al.* Vesta's mineralogical composition as revealed by the visible and infrared spectrometer on Dawn. *Meteorit. Planet. Sci.* **48**, 1–19 <http://dx.doi.org/10.1111/maps.12138> (2013).
34. Li, J.-Y. *et al.* Photometric properties of Vesta. In *Asteroids, Comets, Meteors Conf.* (eds Li, J.-Y. *et al.*) abstr. 6387 (2012).
35. Adams, J. B. Visible and near IR diffuse reflectance spectra of pyroxenes as applied to remote sensing of solid objects in the solar system. *J. Geophys. Res.* **79**, 4829–4836 (1974).
36. Singer, R. B. Near-infrared spectral reflectance of mineral mixtures: systematic combinations of pyroxenes, olivine and iron oxides. *J. Geophys. Res.* **86**, 7967–7982 (1981).
37. Cloutis, E. A. & Gaffey, M. J. Spectral-compositional variations in the constituent minerals of mafic and ultramafic assemblages and remote sensing implications. *Earth Moon Planets* **53**, 11–53 (1991).
38. Sunshine, J. M., Pieters, C. M. & Pratt, S. F. Deconvolution of mineral absorption bands: an improved approach. *J. Geophys. Res.* **95**, 6955–6966 (1990).

39. Sunshine, J. M. & Pieters, C. M. Estimating modal abundances from the spectra of natural and laboratory pyroxene mixtures using the modified Gaussian model. *J. Geophys. Res.* **98**, 9075–9087 (1993).
40. Sunshine, J. M. & Pieters, C. M. Determining the composition of olivine from reflectance spectroscopy. *J. Geophys. Res.* **103**, 13675–13688 (1998).
41. Kanner, L. C., Mustard, J. F. & Gendrin, A. Assessing the limits of the Modified Gaussian Model for remote spectroscopic studies of pyroxenes on Mars. *Icarus* **187**, 442–456 (2007).
42. Clenet, H. Télédétection hyperspectrale: minéralogie et pétrologie, application au volcan Syrtis Major (Mars) et à l'ophiolite d'Oman. PhD thesis, Univ. Toulouse, <http://thesesups.ups-tlse.fr/501/> (2009).
43. Clenet, H., Isaacson, P. J. & Gillet, P. Systematic mapping of mafic minerals in the Copernicus region, the Moon: an improved approach based on Modified Gaussian Model applied to M3 data. *Lunar Planet. Sci. Conf.* **1822** (2014).
44. Clenet, H. *et al.* A systematic mapping procedure based on the Modified Gaussian Model to characterize magmatic units from olivine/pyroxenes mixtures: application to the Syrtis Major volcanic shield on Mars. *J. Geophys. Res.* **118**, 1632–1655 (2013).
45. Beck, P. *et al.* NIR spectral trends of HED meteorites: can we discriminate between the magmatic evolution, mechanical mixing and observation geometry effects? *Icarus* **216**, 560–571 (2011).



Extended Data Figure 1 | Olivine-diogenite spectra and MGM result for NWA4223. Olivine–diogenite spectra from meteorites NWA5480 and NWA4223 (57% and 50% olivine respectively) and comparison with a spectrum from diogenite Tatahouine (orthopyroxenite). Spectra are represented in reflectance space (**a**) and with the continuum removed (**b**) for visual comparison of the shape of the absorption. The MGM result for the

NWA4223 spectrum is represented in **c**. The parameters (from left to right the centre, width and strength are shown) for each Gaussian function (three for olivine and two for orthopyroxene), the continuum and the residual (root mean square, RMS) are reported. Olivine and pyroxene are both correctly detected. Opx, orthopyroxene; Ol, olivine.

Genetics of ecological divergence during speciation

Matthew E. Arnegard^{1,2}, Matthew D. McGee³, Blake Matthews⁴, Kerry B. Marchinko², Gina L. Conte², Sahriar Kabir², Nicole Bedford², Sara Bergek^{5†}, Yingguang Frank Chan⁶, Felicity C. Jones⁶, David M. Kingsley⁶, Catherine L. Peichel¹ & Dolph Schluter²

Ecological differences often evolve early in speciation as divergent natural selection drives adaptation to distinct ecological niches, leading ultimately to reproductive isolation. Although this process is a major generator of biodiversity, its genetic basis is still poorly understood. Here we investigate the genetic architecture of niche differentiation in a sympatric species pair of threespine stickleback fish by mapping the environment-dependent effects of phenotypic traits on hybrid feeding and performance under semi-natural conditions. We show that multiple, unlinked loci act largely additively to determine position along the major niche axis separating these recently diverged species. We also find that functional mismatch between phenotypic traits reduces the growth of some stickleback hybrids beyond that expected from an intermediate phenotype, suggesting a role for epistasis between the underlying genes. This functional mismatch might lead to hybrid incompatibilities that are analogous to those underlying intrinsic reproductive isolation but depend on the ecological context.

The adaptation of populations to contrasting environments is a primary mechanism for the origin of species^{1–4}. In this process, divergent selection leads to high performance of individuals exploiting alternative ecological niches through cumulative changes in potentially many traits⁵. These traits may include morphological phenotypes involved in locomotion and prey capture, behavioural traits that affect encounter rates with different prey types, and phenotypes conferring defence against niche-specific enemies². The complex phenotypic basis of niche use and classic genetic models of adaptation predict that divergence in niche use will have a multilocus genetic architecture with a substantial additive component^{6,7}. However, ecological divergence is often rapid and repeatable and may occur with gene flow⁴, raising the possibility that niche divergence might be accomplished by a few key genomic regions^{8,9}. Although the genetics of putatively adaptive traits have been widely investigated, testing these alternative predictions requires understanding of how genetic changes combine to determine whole-organism performance in different ecological niches^{10,11}.

Because feeding success in different trophic niches depends on an individual's phenotype and environment, we designed a new approach to evaluate predictions about its genetic basis. First we used a semi-natural setting that contained a resource distribution resembling the natural environment and allowed individuals to move freely between trophic niches. We then identified the morphological traits contributing to niche use and feeding performance, and mapped these traits genetically. To confirm that detected loci underlie trophic variation, we fitted the relationship between niche use and genotypes underlying the traits. Finally, we tested the fit of alternative genetic hypotheses of additive, dominance and epistatic effects to axes of feeding variation.

We mapped the genetic basis of niche divergence between the 'benthic' and 'limnetic' species of threespine stickleback fish (*Gasterosteus aculeatus* complex) coexisting in Paxton Lake, British Columbia, Canada.

This pair of species is one of several that evolved independently in postglacial lakes in as few as 12,000 generations by adaptation to alternative niches and frequency-dependent natural selection from resource competition^{12–14}. Benthic and limnetic sticklebacks show nearly complete assortative mating¹⁵ and differ in multiple morphological traits that adapt them to contrasting inshore and pelagic lake habitats, respectively^{14,16–19}. Each species pair probably arose from a double lake invasion from the sea¹², followed by further divergence with gene flow^{16,20}. Hybrids are intermediate in morphology and are outperformed by each parental species in the preferred parental habitats^{14,21–23}. Little intrinsic postzygotic isolation has evolved between the species: laboratory-reared hybrids are viable and fertile^{16,21}.

Niche use and hybrid feeding performance

Just before the breeding season in spring, we introduced 40 F₁ hybrids to an outdoor experimental pond approximating the environmental conditions and contrasting habitats of Paxton Lake (Extended Data Fig. 1 and Supplementary Discussion). We retrieved 633 F₂ hybrid juveniles before their first winter and quantified diet variation among them with the use of stable isotopes ($\delta^{13}\text{C}$ and $\delta^{15}\text{N}$; Fig. 1a). In nature, the use of open water resources by limnetic individuals gives them a lower $\delta^{13}\text{C}$ and higher $\delta^{15}\text{N}$ than the more littoral-feeding benthics, and isotope variation is correlated with foraging trait morphology¹⁷. Body size (length in millimetres) was our measure of F₂ hybrid feeding performance, reflecting how successfully the juveniles acquired food resources and grew during the experiment (Supplementary Discussion). Rapid attainment of adult body sizes often confers fitness advantages to sticklebacks through the effects of size on the avoidance of insect predators²⁴, overwinter survival²⁵, male resource holding potential²⁶ and female fecundity¹⁴.

Under our experimental conditions, the major axis of bivariate isotope variation among F₂ hybrids (principal component 1 (PC1), hereafter

¹Fred Hutchinson Cancer Research Center, Human Biology and Basic Sciences Divisions, 1100 Fairview Avenue North, Seattle, Washington 98109, USA. ²University of British Columbia, Biodiversity Research Centre and Zoology Department, 6270 University Boulevard, Vancouver, British Columbia V6T 1Z4, Canada. ³University of California at Davis, Department of Evolution and Ecology, One Shields Avenue, Davis, California 95616, USA. ⁴EAWAG, Department of Aquatic Ecology, Center for Ecology, Evolution, and Biogeochemistry, Seestrasse 79, 6047 Kastanienbaum, Switzerland. ⁵Uppsala University, Department of Animal Ecology, Evolutionary Biology Centre (EBC), Norbyvägen 18D, SE-75236 Uppsala, Sweden. ⁶Stanford University School of Medicine, Department of Developmental Biology and Howard Hughes Medical Institute, 279 Campus Drive, Stanford, California 94305, USA. [†]Present address: Swedish University of Agricultural Sciences, Department of Aquatic Resources, Stångholmssvägen 2, SE-17893 Drottningholm, Sweden.

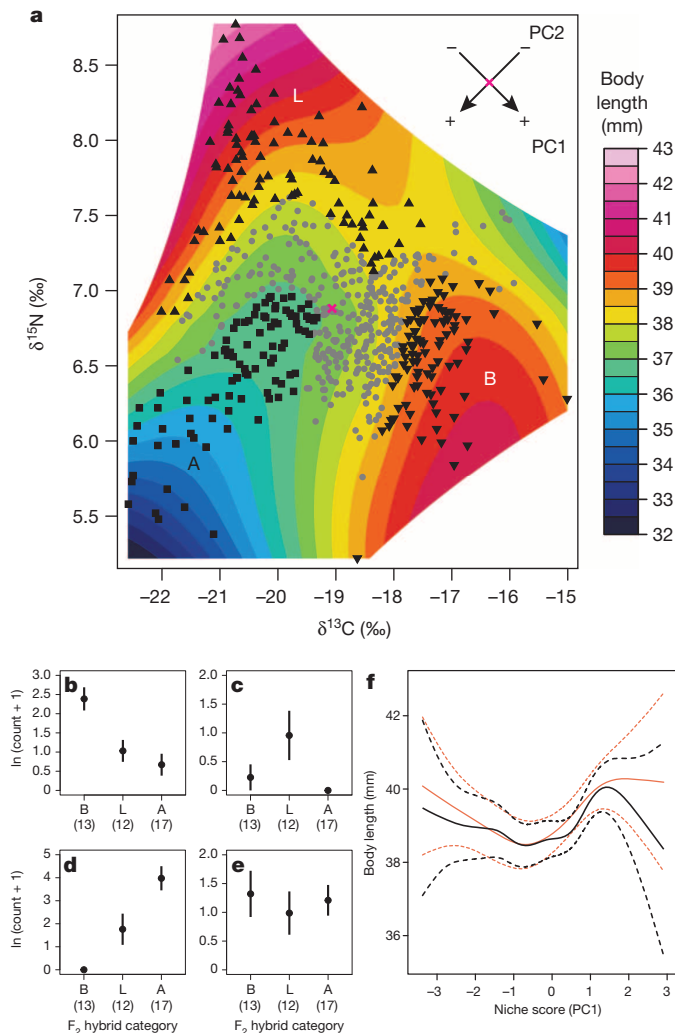


Figure 1 | Niche use and body size. **a**, Stable isotopes ($\delta^{13}\text{C}$ and $\delta^{15}\text{N}$) for 625 F_2 hybrids, showing contours of loess-smoothed body size. Individuals with extreme loess-predicted size are shown as black points (triangles point down, group B; triangles point up, group L; squares, group A; each contains 15% of individuals sampled from the pond; group L restricted to $\text{PC1} < 0.045$ to preserve group distinctiveness). Other individuals are shown as grey circles. Arrows indicate principal components of isotope distribution (PC1 , niche score; PC2 , diet deviation score; origin, red cross). **b–e**, Counts of common food items (means ± 1 s.e.m.) in digestive tracts of group B, L and A individuals. **b**, Larval Chironomidae (benthic macroinvertebrate); **c**, *Skistodiaptomus oregonensis* (evasive calanoid copepod); **d**, Collembola (terrestrial origin, surface dwelling); **e**, *Chydorus* sp. (littoral cladoceran). Kruskal–Wallis tests for differences among groups: larval Chironomidae ($\chi^2_2 = 13.52$, $P = 0.001$); *S. oregonensis* ($\chi^2_2 = 7.547$, $P = 0.023$); Collembola ($\chi^2_2 = 18.67$, $P = 8.82 \times 10^{-5}$); *Chydorus* sp. ($\chi^2_2 = 0.629$, $P = 0.730$). Numbers in parentheses are values of n . **f**, Cubic splines⁴⁸ of mean body size against niche score (predicted values ± 2 s.e.m.) estimated with the 20 largest F_2 families ($n = 438$ individuals), 1,000 bootstrap replicates, and F_2 family as a covariate (black, all individuals; orange, individuals with $\text{PC2} < 0$).

‘niche score’; Fig. 1a) was consistent with the primary axis of limnetic–benthic niche divergence based on isotope data from multiple stickleback species pairs in nature¹⁷ (Supplementary Discussion). A secondary axis of feeding variation (PC2) was also identified. To illustrate variation in phenotype and diet across isotope space, we compared recently consumed prey items among F_2 hybrids from three regions of isotope space (Fig. 1a), which we delineated using loess-predicted body size contours surrounding individuals with the largest (groups L and B) or smallest (A) average body sizes. Individuals in group B had isotope signatures resembling those of the benthic species in nature and consumed significantly

more larval chironomids (Fig. 1b), on which wild benthics specialize^{14,18,19}. In contrast, individuals in group L had a pelagic $\delta^{13}\text{C}$ signature and preyed most heavily on the calanoid copepod *Skistodiaptomus oregonensis* (Fig. 1c), a key planktonic prey item on which limnetics are specialized^{14,17,19}. The small F_2 hybrids in group A fed predominantly on a symphypleonan springtail species (Fig. 1d), which is not a major dietary component of benthics or limnetics in the native lakes^{14,17}. We therefore refer to PC2 as ‘diet deviation score’ because it reflects variation independent of the typical limnetic–benthic feeding axis. The groups did not differ in their consumption of *Chydorus* sp., a littoral cladoceran (Fig. 1e). Additional analyses of consumed prey using all F_2 individuals confirmed these feeding patterns (Extended Data Fig. 2 and Supplementary Discussion).

Analysis of the variation in juvenile size across the entire isotope space revealed a saddle-shaped landscape (Fig. 1a). F_2 hybrids exploiting either the limnetic (group L) or benthic (group B) extremes of the isotope distribution grew more than the other F_2 hybrids, which either had intermediate niche scores and diets or exhibited an alternative feeding pattern (group A). In nature, benthics grow to a larger adult size than limnetics^{14,16}, in part because they differ in the age of sexual maturity²⁷; however, in our experiment, mean body size was similar between the F_2 hybrids in groups B and L (Fig. 1a). This finding might have resulted from sampling the experimental fish as juveniles or from resource abundance differences between the experimental pond and Paxton Lake. The body size valley at intermediate niche scores (Fig. 1a) persists when F_2 family identity is included as a covariate, which controls for variation in F_2 hatching date and hence fish age (Fig. 1f). Considering the 20 largest F_2 families, niche score was reasonably well fitted by a quadratic regression model including the family covariate ($R^2 = 33.2\%$; $F_{21,416} = 9.847$; $P < 2.20 \times 10^{-16}$). Although we found only suggestive evidence for a positive quadratic term in this model (coefficient estimate = 0.173 ± 0.101 s.e.m.; $P = 0.086$; Supplementary Discussion), within-family regression revealed that 16 families individually showed positive quadratic coefficients, indicating that the dip in body size at intermediate niche score is statistically significant ($P = 0.012$; two-sided binomial test). Overall, these results support the hypothesis that F_2 hybrids with an intermediate trophic phenotype suffered a growth disadvantage.

Morphological basis of niche divergence

Many phenotypic traits contribute to niche score variation. To determine this we measured nine functional morphological traits that are important in prey capture and retention, including craniofacial traits affecting the capacity to generate suction pressure, the speed and extent of jaw protrusion, and the retention of ingested prey items (Fig. 2)^{16–19}. We additionally measured the x and y coordinates of 19 morphological landmarks indicating body and head shape²⁸ (Extended Data Fig. 3), which are expected to influence feeding performance. We used all-subsets linear regression to test effects of functional morphological traits and body shape coordinates, separately, on niche score. The best functional trait models (with a difference in Akaike information criterion (ΔAIC ; see Methods) of between 0 and 2) fitting niche score contained terms for three of the five components of the suction feeding index¹⁸, two key oral jaw traits¹⁹ and both gill raker counts^{16,17} (Supplementary Table 1). The best models fitting niche score to body shape contained terms for 22 of 38 landmark coordinates²⁸ (Supplementary Table 2). Hereafter, we consider the traits included in the best models to be ‘component traits’ of niche divergence between Paxton benthics and limnetics.

Genetic architecture of niche divergence

We conducted quantitative trait locus (QTL) mapping on all measured morphological traits and found 76 significant QTLs, including 41 QTLs for 19 of the 29 component traits. The QTLs show small to moderate phenotypic effect sizes (Supplementary Table 3). Component trait QTLs occur on 14 of the 21 linkage groups (LGs) in the threespine stickleback genome²⁹ (Extended Data Fig. 3), suggesting that multiple genetic factors contribute to niche divergence between Paxton benthics and limnetics. Both among LGs and within certain LGs, we find significant

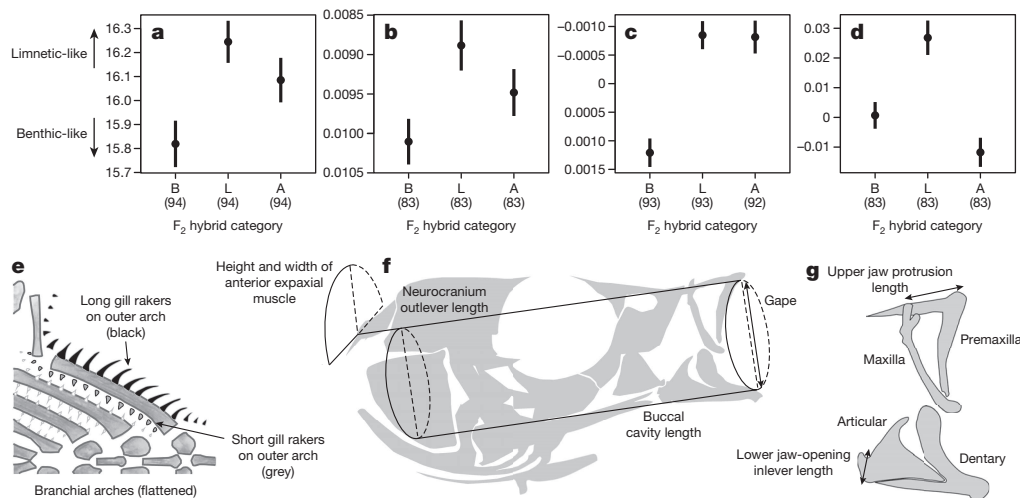


Figure 2 | Trait variation among F₂ hybrid groups. a–d, Trait means (± 1 s.e.m.) of F₂ hybrids in categories B, L and A (Fig. 1a): a, number of short gill rakers (ANOVA, $F_{2,279} = 5.396$, $P = 0.005$); b, suction feeding index ($F_{2,246} = 4.080$, $P = 0.018$); c, residual lower jaw-opening inlever length ($F_{2,275} = 20.36$, $P = 5.65 \times 10^{-9}$); d, residual upper jaw protrusion length ($F_{2,246} = 14.94$, $P = 7.54 \times 10^{-7}$). Numbers in parentheses are values of n .

Arrows indicate directions of benthic–limnetic divergence (vertical axes of b and c are inverted to facilitate visual comparisons). e–g, Trait illustrations: e, gill rakers, functioning in prey retention^{16,17}; f, five components of suction feeding index^{18,19}; g, two oral jaw traits that influence efficiency of capturing evasive zooplankton¹⁹.

clustering of co-localized QTLs (Extended Data Table 1 and Supplementary Discussion), indicating close linkage or pleiotropic effects of genes underlying different component traits of niche use. Nearly all QTLs for the component traits occur in known regions of repeated genomic differentiation between sympatric benthic and limnetic species in multiple lakes³⁰.

To determine how these QTLs contribute to benthic–limnetic niche divergence, we fitted multiple-QTL mapping (MQM) models of niche score to genotypes at QTLs for component traits. We selected only the single morphological QTL with the strongest estimated effect on niche score from each of the 14 LGs containing QTLs for component traits. Although this method is conservative and may underestimate the number of loci underlying niche score, it avoids unduly complex models involving multiple linked loci within LGs. We found additive allelic effects across multiple loci (Fig. 3). Seven of the 14 selected QTLs significantly affected niche score (Extended Data Table 2). Two of these loci resided within clusters of co-localized QTLs on LGs 4 and 16 (Extended Data Fig. 3, Extended Data Table 1 and Supplementary Discussion). However, effect sizes were distributed roughly evenly among the seven significant loci (percentage of total variance explained = 1.16–3.74%; Extended Data Table 2). Next, we allowed all significant pairwise QTL \times QTL interactions to enter the model and followed this by backward elimination of non-significant terms. The resulting ‘full’ MQM model contained four pairwise interactions in addition to main effects representing 11 of the 14 morphological QTLs (Extended Data Table 3).

To test the relative contributions of additive, dominance and pairwise epistatic effects of these loci to niche score, we specified and compared three nested, general linear models at the markers nearest to the 11 significant QTL positions in the full MQM model. The ‘additive’ model contained only the additive effects of the 11 loci (Fig. 3b; adjusted $R^2 = 15.8\%$; $F_{39,484} = 3.514$; $P = 5.76 \times 10^{-11}$; AIC = 1,533.42). By contrast, the ‘additive + dominance’ model contained both additive and dominance effects of these loci (adjusted $R^2 = 14.4\%$; $F_{50,473} = 2.763$; $P = 1.18 \times 10^{-8}$; AIC = 1,551.80). On the basis of AIC, adjusted R^2 , and results of a likelihood ratio (G) test, we conclude that dominance does not contribute significantly to the additive genetic model for niche score ($G_{(\text{add} + \text{dom}, \text{add})} = 3.625$; $P = 0.980$). However, the ‘full’ model, with all additive and dominance effects across the 11 loci as well as the four significant pairwise epistatic interactions (Fig. 3c; adjusted $R^2 = 20.6\%$; $F_{66,457} = 3.059$; $P = 2.75 \times 10^{-12}$; AIC = 1,526.37), provides a significantly better fit to the data than the additive model ($G_{(\text{full}, \text{add})} = 61.05$;

$P = 1.92 \times 10^{-4}$) or the additive + dominance model ($G_{(\text{full}, \text{add} + \text{dom})} = 57.43$; $P = 1.41 \times 10^{-6}$). These results verify the prediction of a polygenic and largely additive basis to whole-organism niche use. F₂ hybrids that grew the most during our study, reflecting high feeding performance, were either those individuals with the highest number of limnetic alleles across loci and the most limnetic-like phenotype and diet, or the highest number of benthic alleles and the most benthic-like phenotype and diet (Figs 1b, c, 2a–d and 3a and Extended Data Figs 2, 4–6). Pairwise genetic interactions also had a significant, although smaller, effect on niche score (compare Fig. 3b with Fig. 3c). This is consistent with a role for epistasis in adaptation, although the importance of epistasis may be underestimated because genetic interactions can be difficult to detect, particularly when they are weak or involve more than two interacting genetic factors^{31,32}.

Trait mismatch reduces growth

Analysis of the secondary axis of isotope variation, the diet deviation score, provided additional evidence for non-additive effects. Inspection of phenotypes of hybrids in group A suggests that this subset of individuals experienced growth deficits (Fig. 1a) due to functional mismatch between certain traits. These group A individuals had distinctly limnetic-like lower jaw-opening inlevers (Fig. 2c, g), which contributed to the rapid jaw opening needed for successful strikes on evasive zooplankton such as *S. oregonensis*¹⁹. Yet they also had reduced, or benthic-like, upper jaw protrusion (Fig. 2d, g), which is expected to decrease the efficiency of zooplanktivory¹⁹. The F₂ hybrids in group A also exhibited mismatches in other combinations of traits (Fig. 2a, Extended Data Figs 4 and 5 and Supplementary Discussion). We predict that these conflicting trait combinations would reduce an individual’s foraging success in both parental habitats, which could explain why these hybrids, as a group, were the smallest of any phenotypic class (Fig. 1a). This phenotypic interaction would imply epistasis for performance at underlying genes even if the phenotypic traits themselves have a largely additive genetic basis³³. Such epistatic effects are expected to be manifested only in environments containing the divergent habitats to which the parental populations are adapted.

To investigate further, we applied the approach used in our genetic analysis of niche score to diet deviation score. Although many morphological traits underlie variation along this secondary feeding axis, MQM modelling revealed no statistically significant relationship between the

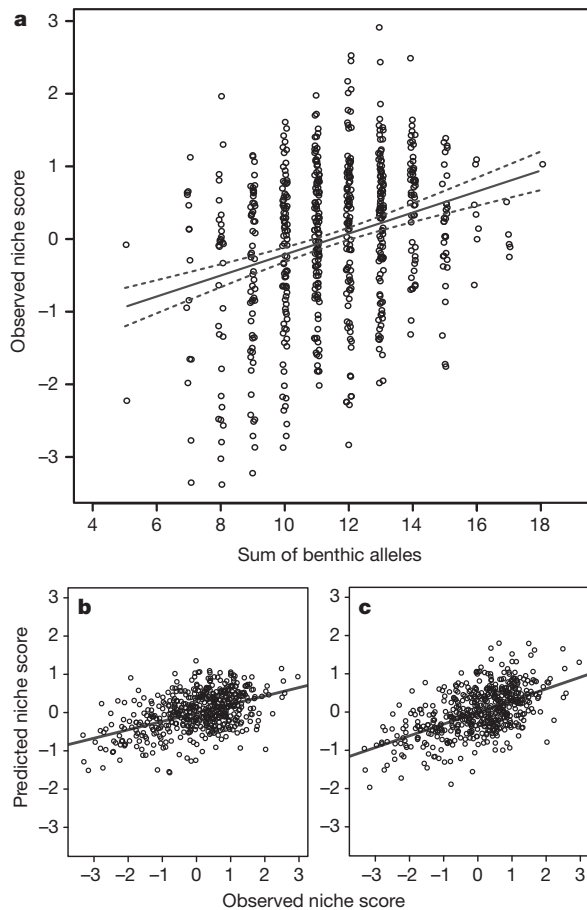


Figure 3 | Genetic architecture of niche divergence. **a**, Niche scores of F_2 hybrids are predicted from the number of benthic alleles summed across 11 unlinked loci in the full MQM model ($R^2 = 0.081$; $F_{1,605} = 53.52$; $P = 8.18 \times 10^{-13}$). Dashed lines are 95% confidence intervals of regression line (solid). **b**, Observed niche score compared with that predicted by the additive-only genetic model. **c**, Observed niche score compared with that predicted by the full genetic model: additive, dominance and epistatic effects. Statistics for **b** and **c** are provided in the text.

QTLs for these traits and diet deviation score. Consequently, we focused on strongest-effect QTLs for the two traits that showed clear phenotypic mismatch in group A individuals (Fig. 2c, d), had strong effects on niche score (Extended Data Fig. 6 and Supplementary Table 1) and were among the most divergent functional morphological phenotypes known in the species pair¹⁹ (that is, the QTLs at 28.8 centimorgans (cM) on LG 4 for lower jaw-opening inlever, and 28.4 cM on LG 9 for upper jaw protrusion; Supplementary Table 3). Using genotypes at the marker nearest to each of these QTLs, we found suggestive evidence of negative epistasis underlying diet deviation score in a two-way analysis of variance (ANOVA) of all F_2 individuals (interaction term: $F_{4,598} = 2.254$; $P = 0.0621$).

Discussion

Although early in the speciation process, Paxton limnetic and benthic sticklebacks differ in many morphological traits. We show that many of these divergent traits contribute to variation in niche use and growth of juvenile F_2 hybrids foraging freely in a semi-natural environment. Multiple genetic factors with largely additive effects, distributed across many chromosomes, underlie niche divergence along the limnetic–benthic resource gradient. Replacement of a limnetic allele by a benthic allele (or vice versa) at any of these loci shifts the niche score in hybrids by roughly the same magnitude (Fig. 3a). We also found evidence for a functional mismatch between phenotypic traits in hybrids that adopted an alternative feeding mode, accompanied by the slowest growth in the

mapping population. This suggests that when multiple traits must function together, novel combinations of traits in hybrids might reduce performance below that expected for an intermediate phenotype. We predict that similar genetic architectures—involving multiple genomic regions each with a relatively small effect, coupled with the possibility of functional mismatch of some gene combinations—will be found for other complex, whole-organism phenotypes that depend on many component traits.

Our finding that niche divergence is determined by small-effect loci on more than half of the chromosomes in the threespine stickleback genome might not be expected in systems in which gene flow still occurs. Theory indicates that loci with relatively large effect sizes under the strongest divergent selection will most readily resist gene flow⁹, and these loci would be detected most readily by QTL mapping. In contrast, loci under weak divergent selection are less able to resist gene flow unless they are sufficiently tightly linked to other loci under sufficiently strong selection^{9,34}. Nevertheless, our results are consistent with genome scans of ecologically divergent populations of several organisms, including threespine sticklebacks, which typically show differences in many regions distributed across the genome^{29,30,35–38}. It is possible that the broadly distributed genetic architecture of niche divergence in the Paxton Lake species pair has arisen from strong, multifarious divergent selection³⁹ acting simultaneously on the numerous traits that underlie adaptation to open-water versus littoral or benthic habitats. Another intriguing possibility is that this broadly distributed genetic architecture results from segregation of ancestral variation that arose during periods of allopatry^{30,38,40}.

Our results contribute to an understanding of the genetics of environment-dependent reproductive isolation during ecological speciation, because divergence in traits underlying niche use reduces the fitness of intermediate phenotypes, including hybrids, when intermediate environments are uncommon or unprofitable^{22,41,42}. Environment-dependent reproductive isolation accompanies the earliest stages of adaptation and may drive the evolution of additional forms of reproductive isolation^{2,43}. If rapid growth of certain threespine stickleback juveniles (such as those in groups L and B) has positive consequences for fitness, then disruptive selection found in the saddle-shaped landscape of body size (Fig. 1a) might reflect selection against intermediate hybrid phenotypes along the major axis of niche differentiation (Fig. 1f). This pattern corroborates the results from transplant experiments in the native lake showing a growth disadvantage in intermediate hybrids relative to the two parental species in their native habitats^{14,21,22}. Thus, our results on the genetics of divergence in niche use and whole-organism performance suggest that the underlying genetic basis of extrinsic postzygotic reproductive isolation between limnetic and benthic sticklebacks is largely additive. This contrasts with the genetics of environment-independent or 'intrinsic' postzygotic isolation, the evolution of which is well described by the Bateson–Dobzhansky–Muller model and is largely caused by negative epistatic interactions between loci^{3,8,44}. Nevertheless, our results show that a mismatch of oral jaw traits reduces feeding performance of some F_2 hybrid sticklebacks beyond that expected from additive genetic effects alone. A functional mismatch between traits might therefore represent an environment-dependent counterpart to the deleterious intermolecular interactions often associated with intrinsic postzygotic isolation⁴⁴. As our results suggest, hybrids that are phenotypically mismatched for ecological performance traits may be produced inevitably as the process of ecological speciation unfolds, thereby contributing to the further evolution of reproductive isolation.

METHODS SUMMARY

We established four F_1 families from crosses between unique wild benthic and limnetic F_0 individuals. On 17 March 2008 we added five adults of each sex and F_1 family to an experimental pond at the University of British Columbia. The 40 F_1 hybrids mated freely to produce a large F_2 intercross population. From 5 to 21 October 2008 we collected and euthanized 633 F_2 juveniles and measured stable carbon and nitrogen isotopes in samples of axial muscle^{17,45}. Prey items were counted in the digestive tracts of 99 of these individuals¹⁴. We analysed body shape with 19

morphometric landmarks placed on images of all fixed and stained individuals²⁸. We also measured nine traits functioning in prey capture or retention^{16–19}. Using all-subsets linear regression, we identified ‘component traits’ that predict variation along principal component 1 (PC1; niche score) or 2 (PC2; diet deviation score) of the stable-isotope distribution. We genotyped F_0 , F_1 and F_2 individuals at 408 single-nucleotide polymorphisms (SNPs)³⁰ and used JoinMap 3.0⁴⁶ to construct a linkage map. Data from 530 individuals in the 29 F_2 families containing at least eight full sibs each were used to interval map traits in R/qtl with Haley–Knott regression and F_2 family as a covariate⁴⁷. Genome-wide LOD significance thresholds ($\alpha = 0.05$) were estimated by permutation (10,000 iterations per trait). One QTL on each LG with the greatest estimated effect on PC1 (or PC2) was identified as the highest-LOD QTL among component traits mapping to that LG. We tested how the identified QTLs affected either trophic axis by fitting MQM models for PC1 or PC2 with Haley–Knott regression and the F_2 family covariate⁴⁷. Additive, dominance and pairwise epistatic effects in the final MQM model for PC1 were tested with nested general linear models and compared using AIC, adjusted R^2 , and likelihood ratio tests.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 June 2013; accepted 1 April 2014.

Published online 8 June 2014.

- Darwin, C. *The Origin of Species by Means of Natural Selection* (John Murray, 1859).
- Schluter, D. *The Ecology of Adaptive Radiation* (Oxford Univ. Press, 2000).
- Coyne, J. A. & Orr, H. A. *Speciation* (Sinauer Associates, 2004).
- Nosil, P. *Ecological Speciation* (Oxford Univ. Press, 2012).
- Chase, J. M. & Leibold, M. A. *Ecological Niches: Linking Classical and Contemporary Approaches* (Univ. of Chicago Press, 2003).
- Fisher, R. A. *The Genetical Theory of Natural Selection* (Oxford Univ. Press, 1930).
- Orr, H. A. The genetic theory of adaptation: a brief history. *Nature Rev. Genet.* **6**, 119–127 (2005).
- Gavrilets, S. *Fitness Landscapes and the Origin of Species* (Monographs in Population Biology Vol. 41, Princeton Univ. Press, 2004).
- Yeaman, S. & Whitlock, M. C. The genetic architecture of adaptation under migration-selection balance. *Evolution* **65**, 1897–1911 (2011).
- Mackay, T. F. C., Stone, E. A. & Ayroles, J. F. The genetics of quantitative traits: challenges and prospects. *Nature Rev. Genet.* **10**, 565–577 (2009).
- Barrett, R. D. H. & Hoekstra, H. E. Molecular spandrels: tests of adaptation at the genetic level. *Nature Rev. Genet.* **12**, 767–780 (2011).
- Taylor, E. B. & McPhail, J. D. Historical contingency and ecological determinism interact to prime speciation in sticklebacks, *Gasterosteus*. *Proc. R. Soc. Lond. B* **267**, 2375–2384 (2000).
- Schluter, D. Experimental evidence that competition promotes divergence in adaptive radiation. *Science* **266**, 798–801 (1994).
- Schluter, D. Adaptive radiation in sticklebacks: trade-offs in feeding performance and growth. *Ecology* **76**, 82–90 (1995).
- Rundle, H. D., Nagel, L. N., Boughman, J. W. & Schluter, D. Natural selection and parallel speciation in sympatric sticklebacks. *Science* **287**, 306–308 (2000).
- McPhail, J. D. Ecology and evolution of sympatric sticklebacks (*Gasterosteus*): evidence for a species pair in Paxton Lake, Texada Island, British Columbia. *Can. J. Zool.* **70**, 361–369 (1992).
- Matthews, B., Marchinko, K. B., Bolnick, D. I. & Mazumder, A. Specialization of trophic position and habitat use by sticklebacks in an adaptive radiation. *Ecology* **91**, 1025–1034 (2010).
- McGee, M. D. & Wainwright, P. C. Convergent evolution as a generator of phenotypic diversity in threespine stickleback. *Evolution* **67**, 1204–1208 (2013).
- McGee, M. D., Schluter, D. & Wainwright, P. C. Functional basis of ecological divergence in sympatric stickleback. *BMC Evol. Biol.* **13**, 277 (2013).
- Gow, J. L., Peichel, C. L. & Taylor, E. B. Contrasting hybridization rates between sympatric three-spined sticklebacks highlight the fragility of reproductive barriers between evolutionarily young species. *Mol. Ecol.* **15**, 739–752 (2006).
- Hatfield, T. & Schluter, D. Ecological speciation in sticklebacks: environment-dependent hybrid fitness. *Evolution* **53**, 866–873 (1999).
- Rundle, H. D. A test of ecologically dependent postmating isolation between sympatric sticklebacks. *Evolution* **56**, 322–329 (2002).
- Gow, J. L., Peichel, C. L. & Taylor, E. B. Ecological selection against hybrids in natural populations of sympatric threespine sticklebacks. *J. Evol. Biol.* **20**, 2173–2180 (2007).
- Marchinko, K. B. Predation's role in repeated phenotypic and genetic divergence of armor in threespine stickleback. *Evolution* **63**, 127–138 (2009).
- Carlson, S. M., Kottas, A. & Mangel, M. Bayesian analysis of size-dependent overwinter mortality from size-frequency distributions. *Ecology* **91**, 1016–1024 (2010).
- Candolin, U. & Voigt, H.-R. Correlation between male size and territory quality: consequence of male competition or predation susceptibility? *Oikos* **95**, 225–230 (2001).
- MacColl, A. D. C. Parasites may contribute to ‘magic trait’ evolution in the adaptive radiation of three-spined sticklebacks, *Gasterosteus aculeatus* (Gasterosteiformes: Gasterosteidae). *Biol. J. Linn. Soc.* **96**, 425–433 (2009).
- Rogers, S. M. *et al.* Genetic signature of adaptive peak shift in threespine stickleback. *Evolution* **66**, 2439–2450 (2012).
- Jones, F. C. *et al.* The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55–61 (2012).
- Jones, F. C. *et al.* A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. *Curr. Biol.* **22**, 83–90 (2012).
- Phillips, P. C. Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Rev. Genet.* **9**, 855–867 (2008).
- Mackay, T. F. C. Epistasis and quantitative traits: using model organisms to study gene–gene interactions. *Nature Rev. Genet.* **15**, 22–33 (2014).
- Whitlock, M. C., Phillips, P. C., Moore, F. B.-G. & Tonsor, S. J. Multiple fitness peaks and epistasis. *Annu. Rev. Ecol. Syst.* **26**, 601–629 (1995).
- Via, S. Natural selection in action during speciation. *Proc. Natl Acad. Sci. USA* **106**, 9939–9946 (2009).
- Hohenlohe, P. A. *et al.* Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet.* **6**, e1000862 (2010).
- Feder, J. L., Egan, S. P. & Nosil, P. The genomics of speciation-with-gene-flow. *Trends Genet.* **28**, 342–350 (2012).
- Strasburg, J. L. *et al.* What can patterns of differentiation across plant genomes tell us about adaptation and speciation? *Phil. Trans. R. Soc. B* **367**, 364–373 (2012).
- Seehausen, O. *et al.* Genomics and the origin of species. *Nature Rev. Genet.* **15**, 176–192 (2014).
- Rice, W. R. & Hostert, E. E. Laboratory experiments on speciation: what have we learned in 40 years? *Evolution* **47**, 1637–1653 (1993).
- Schluter, D. & Conte, G. L. Genetics and ecological speciation. *Proc. Natl Acad. Sci. USA* **106**, 9955–9962 (2009).
- Egan, S. P. & Funk, D. J. Ecologically dependent postmating isolation between sympatric host forms of *Neochlamis bebbianae* leaf beetles. *Proc. Natl Acad. Sci. USA* **106**, 19426–19431 (2009).
- McBride, C. S. & Singer, M. C. Field studies reveal strong postmating isolation between ecologically divergent butterfly populations. *PLoS Biol.* **8**, e1000529 (2010).
- Rundle, H. D. & Nosil, P. Ecological speciation. *Ecol. Lett.* **8**, 336–352 (2005).
- Presgraves, D. C. The molecular evolutionary basis of species formation. *Nature Rev. Genet.* **11**, 175–180 (2010).
- Fry, B. *Stable Isotope Ecology* (Springer, 2006).
- van Ooijen, J. W. & Voorrips, R. E. *JoinMap® 3.0: Software for the Calculation of Genetic Linkage Maps* (Plant Research International, 2001).
- Broman, K. W. & Sen, S. *A Guide to QTL Mapping with R/qtl* (Springer Science+Business Media, 2009).
- Schluter, D. Estimating the form of natural selection on a quantitative trait. *Evolution* **42**, 849–861 (1988).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Perez for counting gill rakers; C. Sather for performing lab work for SNP genotyping; and K. Broman, G. Coop, I. Goodliffe, A. Greenwood, P. Wainwright, M. White and M. Wund for constructive comments. Stable isotopes were analysed at the University of California, Davis, Stable Isotope Facility. Funding was provided by grants from the US National Institutes of Health (F32 GM086125 to M.E.A., R01 GM089733 to C.L.P. and D.S., and P50 HG002568 to C.L.P. and D.M.K.), the Natural Sciences and Engineering Research Council of Canada (to D.S.) and the Canada Foundation for Innovation (to D.S.).

Author Contributions M.E.A., C.L.P. and D.S. designed, planned and oversaw the project. M.E.A. made the crosses, set up the experimental pond and coordinated all field and laboratory research. M.E.A., K.B.M., S.K., N.B. and S.B. conducted fieldwork and stable-isotope analysis. M.D.M. measured functional morphological traits. B.M. and M.E.A. measured and analysed gut contents. S.K., D.S. and M.E.A. performed landmark-based morphometric analyses. M.E.A. analysed relationships between all traits and trophic variation. F.C.J., Y.F.C. and D.M.K. designed the SNP genotyping array. M.E.A., G.L.C., C.L.P. and D.S. analysed SNP genotypes. D.S. determined the genealogy of the mapping population on the basis of SNP genotypes. M.E.A., C.L.P. and D.S. performed linkage and QTL analysis. M.E.A., C.L.P. and D.S. tested the genetic architecture of niche divergence. M.E.A., C.L.P., D.S., M.D.M., B.M. and G.L.C. interpreted the results. M.E.A. wrote the paper with input from C.L.P. and D.S., who are co-senior authors. All other authors provided editorial comments and approved the final version of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.E.A. (marnegar@fhcrc.org).

METHODS

Experimental pond and F₂ hybrid population. We used an outdoor experimental pond at the University of British Columbia (Vancouver, Canada) containing shallow, littoral and deep, open-water habitats (Extended Data Fig. 1 and Supplementary Discussion). Four *in vitro* crosses were made between unique, wild-caught F₀ Paxton benthics and limnetics. F₀ females were benthic for two crosses and limnetic for the other two. After raising the F₁ families in separate aquariums²⁸, we stocked the pond on 17 March 2008 with five F₁ adults for each sex and family ($n = 40$). No food or nutrients were added to the pond after stocking (Supplementary Discussion). During these procedures, fin clips were removed from F₀ and F₁ individuals and stored in 95% ethanol for genetic analysis.

F₁ adults mated freely in the pond to produce an F₂ population. We collected 633 juvenile F₂ individuals in autumn 2008 (5–21 October), when rapid stickleback growth begins to slow⁴⁹ and before any overwintering mortality²⁵. F₂ hybrids were captured with unbaited, fine-mesh minnow traps set in all parts of the pond. During fieldwork we selected 99 F₂ hybrids (in a blind manner) taken from traps deployed for no longer than 2 h. Each of these individuals was euthanized and preserved immediately for subsequent analysis of consumed food items in its digestive tract. All other individuals were housed in tanks and processed within 24 h. F₁ adults were readily excluded by size.

Niche use by F₂ juveniles. We euthanized F₂ hybrids with an overdose of buffered tricaine methanesulfonate and rinsed them in distilled water. Caudal and left pectoral fins were removed and stored in 95% ethanol for genetic analysis. Using a clean scalpel, we sampled white skeletal muscle from the posterior left flank, excluding any skin or bone, and immediately freeze-dried the samples in a BenchTop Manifold Freeze Dryer (Millrock Technology Inc.). Fish were fixed in 7.5% formalin (phosphate-buffered) for 1 month, and then transferred to 40% propan-2-ol.

We homogenized the freeze-dried muscle samples and took 0.8–1.2-mg subsamples, which were enclosed in tin capsules (Elemental Microanalysis Ltd), placed in 96-well microplates and stored in a vacuum-sealed desiccator. The subsamples were assayed for stable isotopes of carbon (¹²C and ¹³C) and nitrogen (¹⁴N and ¹⁵N) at the University of California, Davis, Stable Isotope Facility in one continuous run in January 2009. Measurements were made with a PDZ Europa ANCA-GSL elemental analyser interfaced to a PDZ Europa 20–20 mass spectrometer (Sercon Ltd); these are expressed as scaled isotope ratios, in parts per thousand (‰) relative to Pee Dee Belemnite or atmospheric N₂, using the standard delta notation ($\delta^{13}\text{C}$ or $\delta^{15}\text{N}$)^{45,50}. We performed principal component analysis (PCA) on the bivariate isotope data using the function 'prcomp' in R (v.2.14.0)⁵¹, after scaling both $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ to unit variance. The first PC axis (PC1, niche score) explained 56.5% of total variance in isotope space (first eigenvalue $\lambda_1 = 1.13$); PC2 (diet deviation score) explained the remaining 43.5% of variance ($\lambda_2 = 0.87$).

The $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ signature of skeletal muscle is an integrative measure of an individual's long-term diet (that is, several weeks to months)^{17,45,50,52–57}. We compared these signatures with a direct measure of F₂ hybrid feeding activity immediately (that is, several hours) before capture, which we quantified by means of counts of ingested food items in 99 F₂ hybrids. Food items in the digestive tract of each individual were counted after being sorted into the following 14 categories: adult aquatic snails (class Gastropoda); snail eggs; ostracods (class Ostracoda); calanoid copepods, all identified as *Skistodiaptomus oregonensis* (order Cladocera); cyclopoid copepods (order Cyclopoida); *Chydorus* sp. (order Cladocera); *Sida* sp. (Cladocera); *Gammarus* sp. (order Amphipoda); water mites (unranked taxon Hydrachnidae, suborder Prostigmata); caddisfly larvae (order Trichoptera); chironomid larvae (family Chironomidae); beetle larvae (order Coleoptera); springtails (order Symphyleona, subclass Collembola); and all other terrestrial and surface-dwelling (that is, neustonic) insects, combined. Four categories (chironomids, *S. oregonensis*, springtails and *Chydorus*) accounted for more than 98% of all ingested food items across individuals.

We used body size of fish at capture (length in millimetres) as a measure of feeding performance (Supplementary Discussion). Body size was taken as the distance between morphometric landmarks 1 and 13 (Extended Data Fig. 3). Size variation across the isotope landscape was visualized as the loess (local second-degree polynomial) regression surface of body size on $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$, estimated using the R function 'loess'⁵¹ (span = 0.75). A plot of this surface suggested isotopically distinct regions of extreme performance, reflected by especially large or small average body size of the juveniles in each region (Fig. 1a). To facilitate comparison of diet and morphology among regions, we used contours of the loess regression fit to establish boundaries around individuals of largest or smallest predicted body size. Each boundary was the most extreme predicted size contour enclosing a unique set of individuals numbering about 15% of the distribution ($n = 94$ –95 per region). Thus, region B contained individuals of large average size near the performance peak at high $\delta^{13}\text{C}$ and low $\delta^{15}\text{N}$ (Fig. 1a), whereas region A contained individuals with the smallest average size observed overall, at low $\delta^{13}\text{C}$ and low $\delta^{15}\text{N}$. In these cases, the simple use of appropriate contours allowed a straightforward application

of the 15% criterion. We wanted the third region (L) to also include individuals that grew to large average size (like region B) but were instead located around the performance peak at low $\delta^{13}\text{C}$ and high $\delta^{15}\text{N}$. With region L, however, a second criterion (minimization of PC1) was required to define a boundary that both contained an outer quantile (15%) of the predicted performance distribution and retained isotopic distinctiveness from other regions. Specifically, the boundary of region L was the maximum loess-predicted size contour enclosing 15% of the distribution (around the low- $\delta^{13}\text{C}$ –high- $\delta^{15}\text{N}$ peak) after limiting this region to PC1 < 0.045. Next, we investigated variation in recent feeding activity among these categories of F₂ hybrids with Kruskal–Wallis tests (R function 'kruskal.test'⁵¹) for differences in counts of ingested food items (Fig. 1b–e).

To test the robustness of the performance valley at intermediate niche score (Fig. 1a), we fitted body size to niche score with a cubic spline function including F₂ family identity (indicating the offspring of each unique F₁ × F₁ pairing) as a covariate. Doing so accounts for among-family variation in F₂ age at capture due to variable F₁ breeding times, assuming that unique F₁ pairs mated only once. Supporting this assumption, we found no deviations from unimodal size frequency distributions in F₂ families, judged by visual inspection and Hartigan's dip test⁵⁸ (R package 'dipTest'⁵⁹; 2,000 replicates per Monte Carlo simulation; each $P > 0.175$). Thus, cubic splines were estimated in 'glms' v.4.0 (<http://www.zoology.ubc.ca/~schluter/wordpress/software/#spline>)⁴⁸, using the 20 largest F₂ families (full sibs per family: $n = 12$ –48). Using the best smoothing parameter (that is, λ with lowest cross-validation score), we obtained standard errors of predicted body sizes (1,000 bootstrap replicates). We also evaluated the robustness of the performance valley by quadratic regression of body size on niche score, again using the 20 largest families and the family identity covariate (Supplementary Discussion). The regression was repeated using only individuals for which PC2 < 0 to ensure that presence of the size valley did not depend solely on unusually small individuals with PC2 ≥ 0, including those in region A.

Morphological trait measurements. Three classes of morphological traits known to differ between wild Paxton benthics and limnetics (Supplementary Discussion) were measured: first, morphometric traits reflecting body shape; second, defensive armour traits; and third, single or composite functional traits (head and jaw) with described roles in feeding^{16–19,60}. We measured shape by using the geometric morphometric approach of previous studies of sticklebacks^{28,61}. Fixed specimens were stained for 48 h in 1% aqueous KOH with 0.005% w/v Alizarin Red S (Merck KGaA) and destained in 40% propan-2-ol. A Nikon D1H camera and three strobe lights were used to make a digital image of the right side of each specimen alongside a ruler. We recorded the x and y coordinates of 19 morphological landmarks from these images with 'tpsDig' v.2.12 (ref. 62) (Extended Data Fig. 3). We scaled, rotated and superimposed landmark configurations using Generalized Procrustes analysis⁶³ (R package 'shapes'⁶⁴), after which we used a standard approach to correct for a specimen bending artefact caused by fixation^{28,61,65} (Supplementary Discussion). The resulting x and y coordinates were treated as individual traits when analysing relationships between shape and stable isotopes and performing QTL mapping.

Images enabled the use of a simple ordinal scale for the rapid characterization of three armour traits: pelvic girdle (right side of body) and first and second dorsal spines. These traits received a score of 0 when absent, 2 when present, and 1 when expressed at an intermediate size between these two categories. 'Well-developed' lateral bony plates¹⁶ were also counted along the right flank (that is, any plate whose height was judged to be at least one-third of the individual's body depth at that plate).

We measured functional morphological traits by using methods previously applied to sticklebacks^{18,19,66}. Gill rakers on the left outer branchial arch were counted under a dissecting microscope after removal of the arch and associated cartilage from the opercular cavity. Any stained protuberance was counted as either a long or short gill raker according to position (Fig. 2e). After clearing specimens by immersion in 30% w/v sodium borate with 1% w/v trypsin until translucent, we measured five component traits of the suction feeding index^{18,19} (Fig. 2f): anterior epaxial muscle height (E_H) and width (E_W), neurocranium outlever length (N_{OL}), buccal cavity length (B_L) and gape (G). Suction index was calculated as $(E_W E_H^2)/(3B_L G(N_{OL} - \frac{1}{2}B_L))$. Last, we measured upper jaw (premaxillary) protrusion length and lower jaw-opening inlever length (Fig. 2g)¹⁹.

All functional morphological traits were corrected for body size (length) except long and short gill raker counts, which were uncorrelated with size. We used standardized major axis regression (function 'sma' in R package 'smatr'⁶⁷) to test for differences in allometric scaling relationships of these traits between F₂ hybrids in the mapping population and wild Paxton benthics and limnetics. This revealed no evidence of allometric differences between the experimental fish and natural populations (likelihood ratio tests, 2 d.f. each: $0.09 < P < 0.56$). The traits were therefore size-corrected by expressing them as residuals from ordinary least-squares regression of each trait on body size⁶⁸.

Identifying 'component traits' of niche use. To determine which morphological traits predict variation along the primary trophic axis, we first regressed niche score on each trait separately by simple linear (least-squares) regression. Armour traits were excluded because we had no a priori evidence of specific influences on trophic variation (Supplementary Discussion). Similarly, we excluded suction feeding index, because each of its component traits was being considered. Scatterplots indicated that the data conformed reasonably well to parametric statistical assumptions. Only traits from the significant univariate regression models were considered further. Taking all such traits to be candidate explanatory variables, we performed all-subsets (multiple linear) regression⁶⁹ of niche score on the candidate traits, using the R function 'leaps'⁷⁰. This function returns and orders the best models based on Mallows's C_p (ref. 71), which we converted to the Akaike information criterion (AIC)⁷². Because of partial redundancy between some of the functional traits and craniofacial landmarks, we considered functional morphology and shape trait classes separately when performing these exhaustive searches for trait subsets that best predicted niche score.

The difference between AIC scores of the 'best' (AIC = 0) and 'ith best' models is denoted ΔAIC . We considered all models with $\Delta AIC \leq 2$ to be statistically indistinguishable from the overall 'best' model identified for given class of traits⁷³. Consequently, the full suite of morphological traits for which this approach found similarly strong within-class evidence of an effect on niche score was the union of explanatory variables among all well-supported models ($0 < AIC \leq 2$) across the two trait classes (functional morphology and shape; Supplementary Tables 1 and 2, respectively). Hereafter we refer to this full suite of traits as the component traits of niche use. We repeated this entire procedure for identifying the morphological traits that influence a trophic axis, using diet deviation score as the response variable instead of niche score.

Genotyping and pedigree analysis. We isolated genomic DNA from fin tissue samples taken from the eight F_0 founders, 40 F_1 adults and 633 F_2 juveniles, using digestion with Proteinase K, extraction with phenol-chloroform and precipitation with ethanol⁷⁴. We resuspended DNA in 30 μ l of TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0) and diluted an aliquot of each sample to a concentration of about 25 ng μ l⁻¹ based on PicoGreen assay (Life Technologies). All F_0 and F_1 individuals and 616 F_2 juveniles were genotyped at 408 SNP markers³⁰, which are distributed across the *G. aculeatus* genome and were polymorphic in our mapping population (Supplementary Table 4). Genotyping was performed with Illumina's GoldenGate assay at the Fred Hutchinson Cancer Research Center (Seattle, Washington, USA), using GenomeStudio software (Illumina Inc.) to score genotypes.

We used a Bayesian parentage assignment algorithm⁷⁵ (R package 'MasterBayes'⁷⁶) and all SNP genotypes to estimate the F_1 parentage of every F_2 individual. Posterior probabilities of correct assignments of F_2 hybrids to their estimated pair of F_1 parents were high (mean \pm s.d. 0.999 ± 0.020). Assignments of F_1 hybrids to known F_0 parents were verified (posterior probability = 1 in every case). Using a custom algorithm written in R, we then coded the SNP genotypes for linkage analysis and QTL mapping based on the reconstructed pedigrees for the F_2 hybrids.

Linkage analysis and QTL mapping. Among 728 F_2 hybrids collected in total from the experimental pond ($n = 633$ juveniles in this study; $n = 95$ adult males collected in spring 2009), we used all 594 genotyped individuals in F_2 families with at least ten full sibs (range 10–53 sibs per family) to construct a linkage map. This was done by using JoinMap v.3.0 ('cross pollinator' population code)⁴⁶. All obtainable pairwise (between-SNP) recombination frequencies and associated \log_{10} of odds (LOD) scores were computed separately for each F_2 family. We created a single pairwise data file by concatenating recombination frequencies and LOD scores across families and used this to produce the map (Supplementary Table 4).

We performed QTL mapping on all measured traits in R/qtl⁴⁷, using all F_2 families from the linkage analysis. Retaining all families after excluding F_2 hybrids collected in spring 2009 required a reduction in minimal acceptable family size (to eight full sibs). Accordingly, our data set for QTL mapping consisted of 530 F_2 hybrids in 29 F_2 families (range: 8–48 sibs per family). Using R/qtl function 'scanone' we performed interval mapping on each trait with Haley–Knott regression and F_2 family identity as a covariate. We conducted 10,000 permutations per trait to determine the genome-wide LOD threshold for significant QTLs at $\alpha = 0.05$ (ref. 47). The resulting LOD thresholds ranged from 3.51 to 3.88 across traits (mean 3.63). For every QTL, we estimated the position of the peak LOD score in centimorgans (cM) with a 1.5-LOD confidence interval around the peak⁷⁷. R/qtl function 'fitqtl' was used to estimate the percentage of phenotypic variance explained by each QTL, and 'find.marker'⁴⁷ was used to identify the nearest SNP.

Genetic architecture of niche divergence. We investigated effects of the discovered morphological QTLs on niche divergence between Paxton benthics and limnetics as follows. First, we considered only QTLs underlying component traits of niche use. From these QTLs we selected the single QTL per LG with the highest LOD score among niche use component traits mapping to that LG. This procedure identified 14 candidate morphological QTLs (on different LGs) with hypothesized

genetic effects on niche score. Repeating this procedure for 'diet deviation score', we identified 15 QTLs on different LGs with hypothesized effects on this secondary trophic axis.

To model cumulative effects of the 14 candidate QTLs on the niche score of F_2 hybrids, we specified candidate QTL positions by using R/qtl function 'makeqtl'. We then used 'fitqtl' to fit a MQM model of the main effects of the QTLs on niche score (Extended Data Table 2). Next, we found all significant pairwise QTL \times QTL interactions among the candidate loci by applying the 'addint' function⁴⁷ to the 14 candidate QTLs. We added these interactions to the main-effects-only MQM model and performed backwards stepwise elimination of non-significant terms until arriving at the full MQM model (Extended Data Table 3). At every step, 'fitqtl' was used for model fitting with the Haley–Knott method and the F_2 family covariate.

We repeated this modelling procedure for diet deviation score and its 15 candidate QTLs. In this case we found nine significant QTL \times QTL interactions by using 'addint', but these interactions were not accompanied by significant main effects (data not shown). Consequently, all further model comparisons were focused on testing genetic effects on niche score.

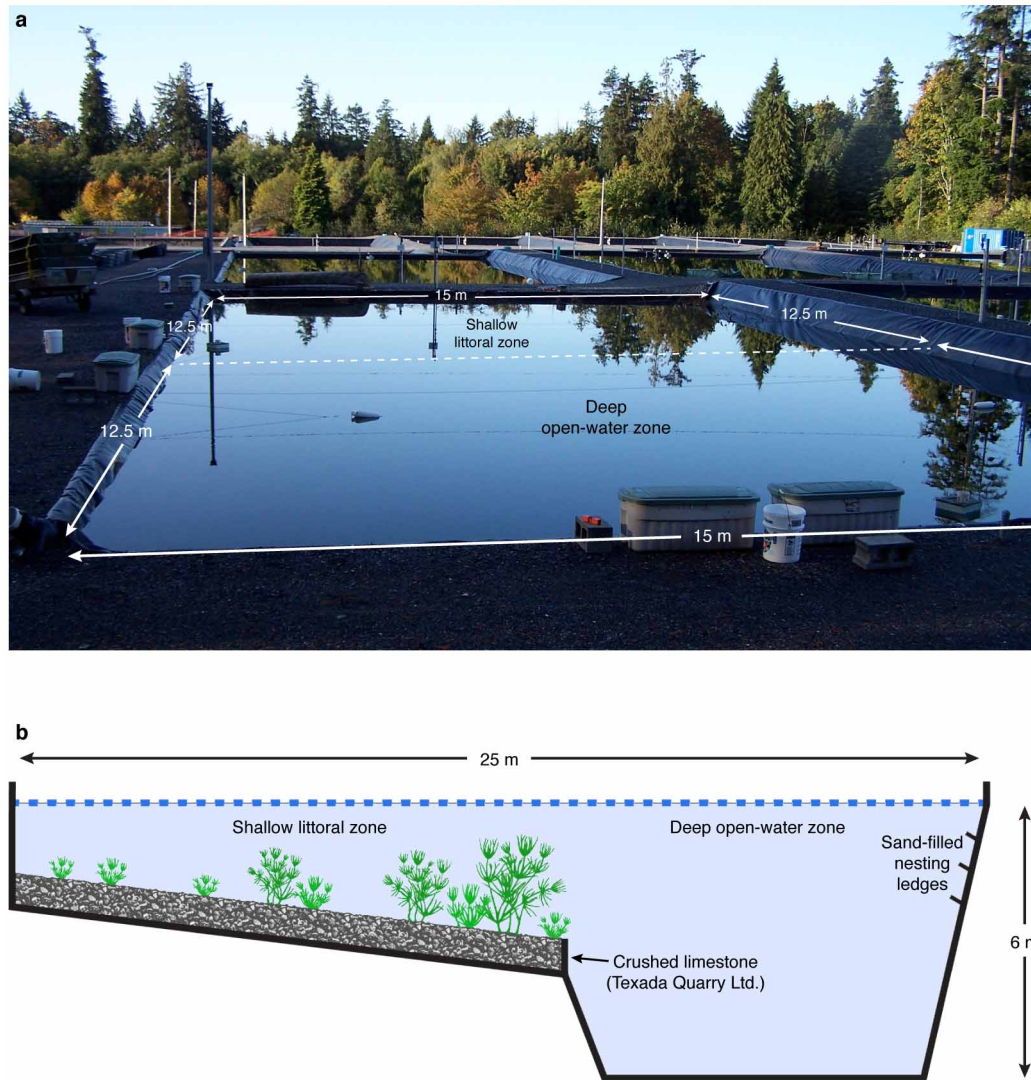
Using the full MQM model, we tested the importance of additive, dominance and pairwise epistatic effects in the genetic architecture of niche divergence between Paxton benthics and limnetics. In R/qtl we imputed genotypes at the SNP marker nearest to each QTL in this model using the Kosambi mapping function⁷⁸. Subsequent model comparisons of QTL effects were performed with the linear model-fitting function 'lm'⁵¹ in R. We specified the 'full' genetic model by coding genotypes as categorical data and including all additive, dominance and pairwise epistatic effects detected in R/qtl for the full MQM model (Extended Data Table 3). Using this genotype coding scheme, an 'additive + dominance' model was specified by including only genotype main effects from the 'full' model. In contrast, an 'additive' model was specified by coding genotypes in terms of the integer number of benthic alleles. Each model again included the F_2 family covariate. Models were compared by using AIC and adjusted R^2 , which penalize models with excessive numbers of terms^{72,73,79,80}, and by using likelihood ratio tests (R function 'ltest'⁸¹).

Animal care, sample size determination and data blinding. All field and laboratory procedures were approved by the University of British Columbia Animal Care Committee (protocols A07-0293 and A11-0402) and the Fred Hutchinson Cancer Research Center Institutional Animal Care and Use Committee (protocol 1797). The target sample size of F_2 hybrids was determined to minimize bias when detecting QTLs (the 'Beavis effect') and to reduce sampling error for estimated QTL effect sizes. Realized sample sizes for QTL mapping ($n = 473$ –530 F_2 hybrids) were sufficient to minimize QTL detection bias and sampling error for effect sizes for every trait considered^{47,82}. All reductions in sample size (from $n = 633$ juveniles collected) occurred in an unbiased fashion, because sample exclusion was based solely on missing phenotype or genotype data, or having too few full sibs in the collection. To avoid sampling and measurement biases, sample identities were not revealed to the authors and technicians who performed phenotypic measurements or genotyping until after all data collection had been completed. The 99 F_2 hybrids allocated for analysis of consumed food items were also selected in a blind manner during fieldwork.

Constructed to help clarify how the two-dimensional isotope distribution of F_2 hybrids related to other patterns of variation, F_2 groups B, L and A each contained about 15% of all F_2 hybrids collected. With the application of a second criterion minimizing PC1 scores in group L only, the 15% inclusion criterion yielded the largest number of individuals in each group without compromising group distinctiveness. Patterns of feeding and morphological variation among groups were robust to alternative body size thresholds (Supplementary Discussion), and they were confirmed by analyses using all individuals (Extended Data Fig. 2 and 6). Moreover, all available data were used for stable-isotope PCA, component trait determination, QTL analysis and genetic modelling; results of these analyses therefore did not depend on how F_2 hybrids were categorized to understand the performance landscape.

49. Wootton, R. J. *A Functional Biology of Sticklebacks* (Univ. of California Press, 1984).
50. Post, D. M. Using stable isotopes to estimate trophic position: models, methods, and assumptions. *Ecology* **83**, 703–718 (2002).
51. R. Development Core Team. *R: A Language and Environment for Statistical Computing* (<http://www.R-project.org/>) (R Foundation for Statistical Computing, 2011).
52. Vander Zanden, M. J. & Vadeboncoeur, Y. Fishes as integrators of benthic and pelagic food webs in lakes. *Ecology* **83**, 2152–2161 (2002).
53. Bolnick, D. I. *et al.* The ecology of individuals: incidence and implications of individual specialization. *Am. Nat.* **161**, 1–28 (2003).
54. McIntyre, P. B. & Flecker, A. S. Rapid turnover of tissue nitrogen of primary consumers in tropical freshwaters. *Oecologia* **148**, 12–21 (2006).
55. Harmon, L. J. *et al.* Evolutionary diversification in stickleback affects ecosystem functioning. *Nature* **458**, 1167–1170 (2009).
56. Behm, J. E., Ives, A. R. & Boughman, J. W. Breakdown in postmating isolation and the collapse of a species pair through hybridization. *Am. Nat.* **175**, 11–26 (2010).

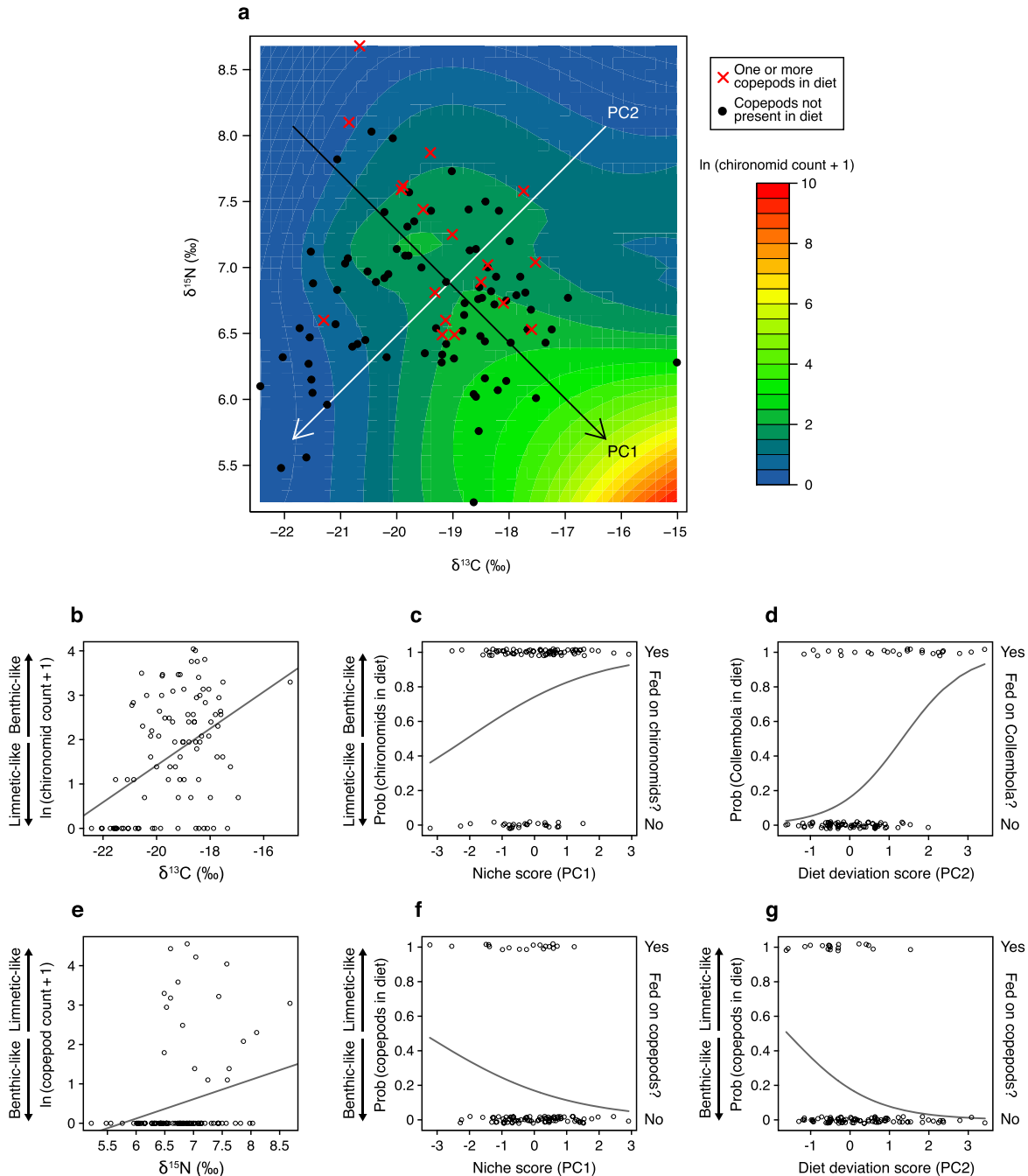
57. Bolnick, D. I. Sympatric speciation in threespine stickleback: why not? *Int. J. Ecol.* **2011**, 942847 (2011).
58. Hartigan, J. A. & Hartigan, P. M. The dip test of unimodality. *Ann. Stat.* **13**, 70–84 (1985).
59. Maechler, M. Package 'dipTest': Hartigan's Dip Test Statistic for Unimodality, corrected code, v. 0.75-4 (<http://CRAN.R-project.org/package=dipTest>) (CRAN: Comprehensive R Archive Network, 2011).
60. Schluter, D. Adaptive radiation in sticklebacks: size, shape, and habitat use efficiency. *Ecology* **74**, 699–709 (1993).
61. Albert, A. Y. K. *et al.* The genetics of adaptive shape shift in stickleback: pleiotropy and effect size. *Evolution* **62**, 76–85 (2008).
62. Rohlf, F. J. *TpsDig2* (<http://life.bio.sunysb.edu/morph/soft-dataacq.html>) (Department of Ecology and Evolution, State Univ. of New York, 2006).
63. Rohlf, F. J. & Slice, D. Extensions of the Procrustes method for the optimal superimposition of landmarks. *Syst. Zool.* **39**, 40–59 (1990).
64. Dryden, I. L. Package 'shapes': Statistical Shape Analysis, v. 1.1-3 (<http://CRAN.R-project.org/package=shapes>) (CRAN: Comprehensive R Archive Network, 2009).
65. Arnegard, M. E. *et al.* Sexual signal evolution outpaces ecological divergence during electric fish species radiation. *Am. Nat.* **176**, 335–356 (2010).
66. Peichel, C. L. *et al.* The genetic architecture of divergence between threespine stickleback species. *Nature* **414**, 901–905 (2001).
67. Warton, D. I., Duursma, R. A., Falster, D. S. & Taskinen, S. SMATR 3—an R package for estimation and inference about allometric lines. *Methods Ecol. Evol.* **3**, 257–259 (2012).
68. Smith, R. J. Use and misuse of the reduced major axis for line-fitting. *Am. J. Phys. Anthropol.* **140**, 476–486 (2009).
69. Miller, A. *Subset Selection in Regression*, 2nd edn, Vol. 95 (Chapman & Hall/CRC, 2002).
70. Lumley, T. Package 'leaps': Regression Subset Selection, v. 2.9 (<http://CRAN.R-project.org/package=leaps>) (CRAN: Comprehensive R Archive Network, 2009).
71. Mallows, C. L. Some comments on C_p . *Technometrics* **15**, 661–675 (1973).
72. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **19**, 716–723 (1974).
73. Burnham, K. P. & Anderson, D. R. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd edn (Springer, 2002).
74. Sambrook, J. & Russell, D. W. *Molecular Cloning: A Laboratory Manual*, 3rd edn (Cold Spring Harbor Laboratory Press, 2001).
75. Hadfield, J. D., Richardson, D. S. & Burke, T. Towards unbiased parentage assignment: combining genetic, behavioural and spatial data in a Bayesian framework. *Mol. Ecol.* **15**, 3715–3730 (2006).
76. Hadfield, J. Package 'MasterBayes': ML and MCMC Methods for Pedigree Reconstruction and Analysis, v. 2.50 (<http://cran.r-project.org/web/packages/MasterBayes/>) (CRAN: Comprehensive R Archive Network, 2012).
77. Manichaikul, A., Dupuis, J., Sen, S. & Broman, K. W. Poor performance of bootstrap confidence intervals for the location of a quantitative trait locus. *Genetics* **174**, 481–489 (2006).
78. Kosambi, D. D. The estimation of map distance from recombination values. *Ann. Eugen.* **12**, 172–175 (1944).
79. Theil, H. *Economic Forecasts and Policy*, 2nd edn (North-Holland, 1961).
80. Engle, R. F. & Brown, S. J. Model selection for forecasting. *Appl. Math. Comput.* **20**, 313–327 (1986).
81. Hothorn, T. *et al.* Package 'lmtest': Testing Linear Regression Models, v. 0.9-30 (<http://CRAN.R-project.org/package=lmtest>) (CRAN: Comprehensive R Archive Network, 2012).
82. Erickson, D. L., Fenster, C. B., Stenøien, H. K. & Price, D. Quantitative trait locus analyses and the study of evolutionary process. *Mol. Ecol.* **13**, 2505–2522 (2004).



Extended Data Figure 1 | Experimental pond used in the study.

a, Photograph of pond no. 4 at the experimental pond facility of the University of British Columbia (Vancouver, British Columbia, Canada), taken in autumn

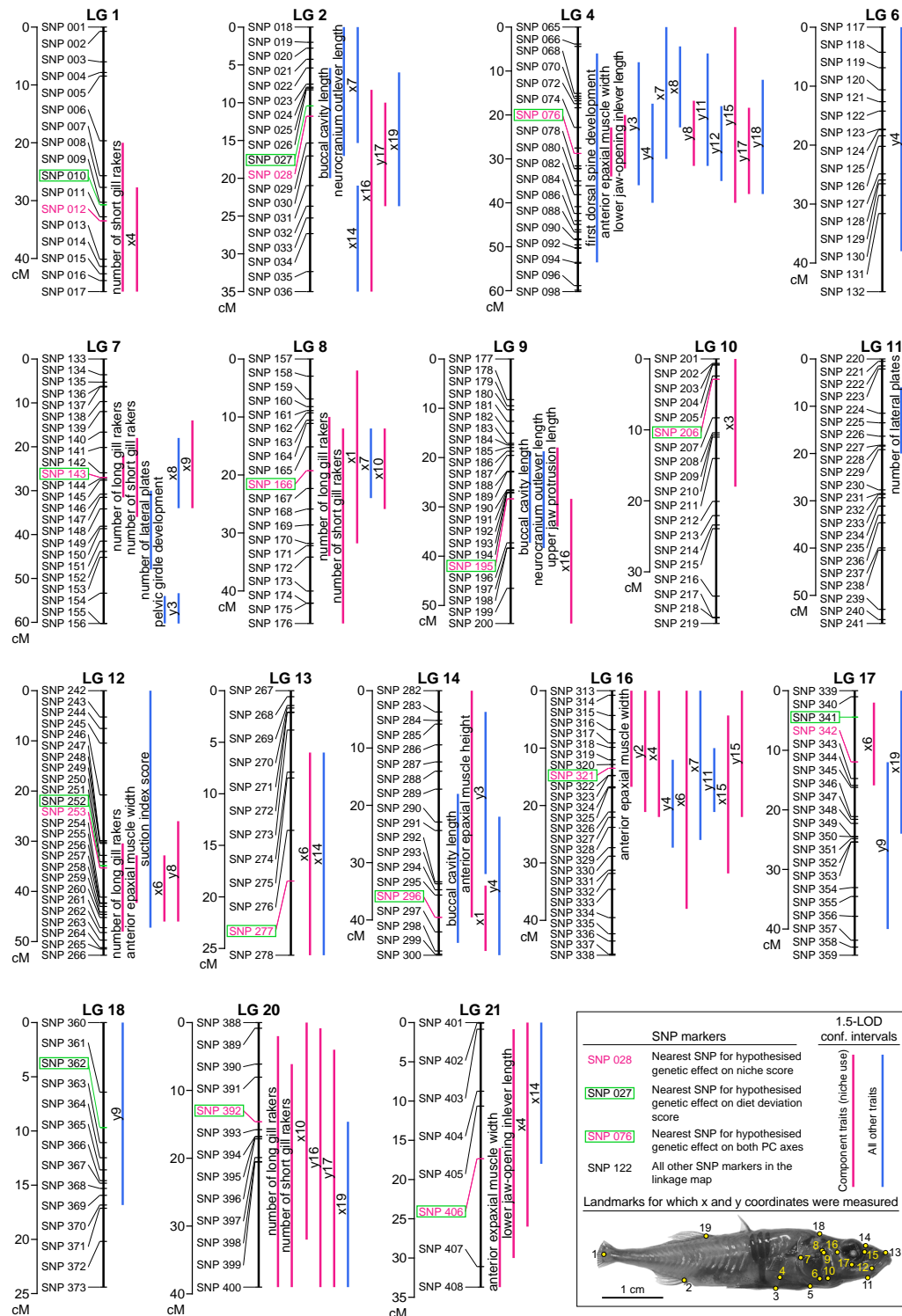
2008, during the collection of F_2 juveniles. **b**, Diagram of the pond profile. See Supplementary Discussion for details on pond history before this study.



Extended Data Figure 2 | Feeding patterns in relation to isotope signatures.

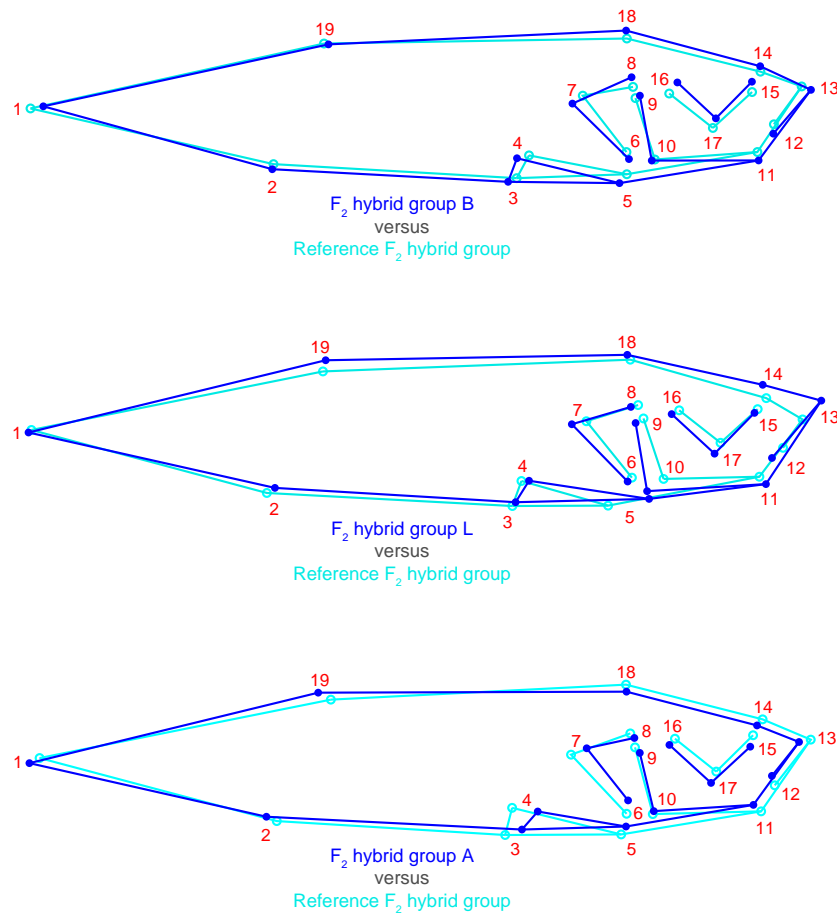
Plots show relationships between ingested prey counts from all available F_2 hybrids ($n = 99$) and stable-isotope data. **a**, Loess-smoothed surface (span = 0.75, second-degree polynomials) of predicted chironomid counts plotted on original isotope axes ($\delta^{13}\text{C}$, $\delta^{15}\text{N}$). As with all other count data plotted here, counts were transformed as $\ln(\text{chironomids} + 1)$ and mapped according to the coloured scale. PC1 (black arrow) and PC2 (white) are based on the entire isotope distribution (Fig. 1a). Individuals are plotted as points according to the presence (crosses) or absence (filled circles) of calanoid copepods in their digestive tracts. **b–g**, Linear or logistic regression, accordingly, of ingested prey count or presence/absence data (transformed

as above) on the different axes through isotope space. **b**, Chironomid count against $\delta^{13}\text{C}$, linear regression, slope estimate = 0.415, $R^2 = 0.199$, $F_{1,97} = 24.1$, $P = 3.70 \times 10^{-6}$. **c**, Chironomid presence against niche score, logistic regression, slope coefficient = 0.504, $z = 2.23$, $P = 0.0255$. **d**, Collembola presence against diet deviation score, logistic regression, slope coefficient = 1.25, $z = 4.26$, $P = 2.03 \times 10^{-5}$. **e**, Calanoid copepod count against $\delta^{15}\text{N}$, linear regression, slope estimate = 0.492, $R^2 = 0.0608$, $F_{1,97} = 6.28$, $P = 0.0139$. **f**, Calanoid copepod presence against niche score, logistic regression, slope coefficient = -0.463 , $z = -1.84$, $P = 0.0651$. **g**, Calanoid copepod presence against diet deviation score, logistic regression, slope coefficient = -0.958 , $z = -2.67$, $P = 0.00766$.



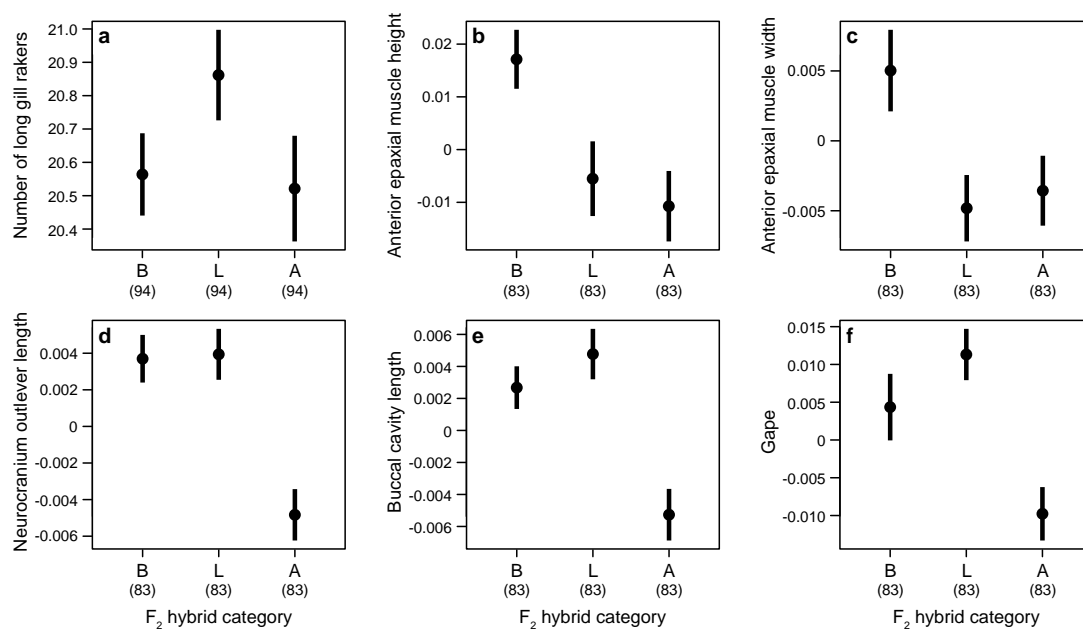
Extended Data Figure 3 | Linkage map showing QTLs for all traits. All *G. aculeatus* chromosomes are represented by LGs in the complete linkage map for this study (LGs and chromosomes use the same numbering²⁹; LGs with no mapped QTLs are omitted here). Map distances are indicated with a scale at the left of each LG in centimorgans (cM). Coloured bars (at the right) are 1.5-LOD confidence intervals for QTL position (red bars, component traits of niche use; blue bars, other traits; Supplementary Table 3 provides LOD scores, map positions of LOD peaks, and effect sizes). The given SNP identifiers (IDs) are only for reference to Supplementary Table 4, which provides published SNP data³⁰. For clarity, every other ID is omitted for SNP 066–098, even though these markers are present in the map. Markers closest to candidate QTLs for genetic model comparisons are highlighted: red text, nearest to candidate QTLs for niche score; green boxes, diet deviation score. Numbered traits are the x and

y coordinates of morphometric landmarks (indicated on the fish photo): 1, posterior midpoint caudal peduncle; 2, anterior insertion anal fin at first soft ray; 3, posteroventral corner ectocoracoid; 4, posterodorsal corner ectocoracoid; 5, anteriormost corner ectocoracoid; 6, anteroventral corner opercle; 7, posterodorsal corner opercle; 8, dorsal edge opercle–hyomandibular boundary; 9, dorsalmost extent preopercle; 10, posteroventral corner preopercle; 11, anteriormost extent preopercle along ventral silhouette; 12, posteroventral extent maxilla; 13, anterodorsal extent maxilla; 14, suture between nasal and frontal bones along dorsal silhouette; 15, anterior margin orbit; 16, posterior margin orbit; 17, ventral margin orbit (landmarks 15–17 placed in line with vertical or horizontal midpoint of eye); 18, posterior extent supraoccipital along dorsal silhouette; 19, anterior insertion dorsal fin at first soft ray.



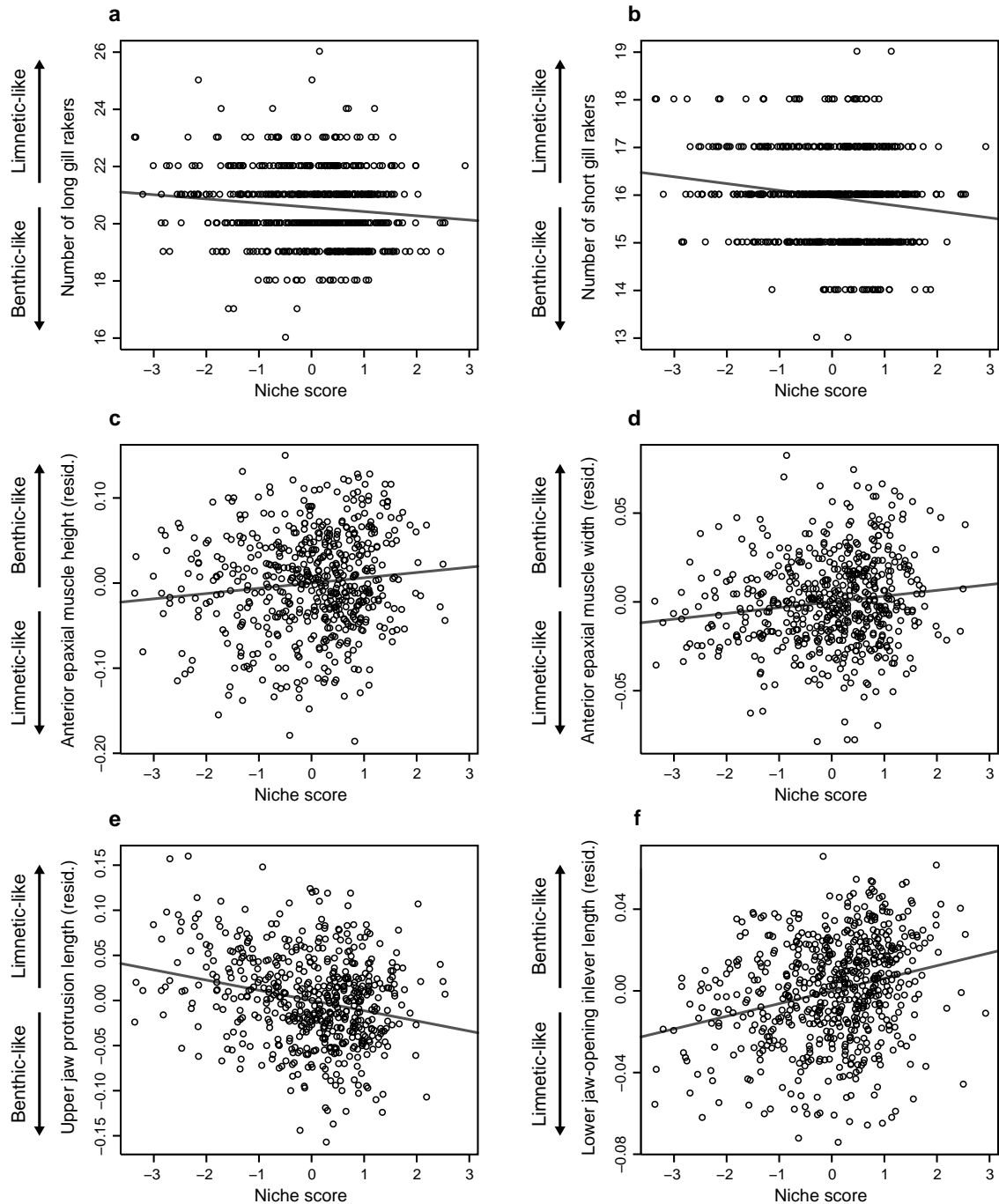
Extended Data Figure 4 | Shape variation among F₂ hybrid groups. Each overlaid pair of wireframe diagrams compares the mean body shape of individuals in one of three groups of F₂ hybrids (B, L or A; shown in dark blue) with the relative mean shape of a reference group consisting of all other F₂ hybrids (group membership shown in Fig. 1a). Using data for 19 Procrustes-superimposed and unbent landmarks (Extended Data Fig. 3), the wireframe diagrams were produced and plotted in MorphoJ v.1.04a, on the basis of

discriminant function analysis (Supplementary Discussion). The shape differences represented here are magnified eightfold for easier visual comparison. Group sample sizes: $n = 91$ (B), $n = 92$ (L), $n = 93$ (A), $n = 335$ (reference group). See Supplementary Discussion for a detailed description of patterns of variation in several specific features of shape that can be interpreted from these data.



Extended Data Figure 5 | Variation of additional traits among F₂ hybrid groups. Means (± 1 s.e.m.) of F₂ hybrids in groups B, L and A (Fig. 1a) are shown for the following traits (using raw data for long gill rakers and size-corrected data for the other traits): **a**, number of long gill rakers (ANOVA, $F_{2,279} = 1.756$, $P = 0.175$); **b**, residual anterior epaxial muscle height ($F_{2,246} = 5.219$, $P = 0.00603$); **c**, residual anterior epaxial muscle width ($F_{2,246} = 4.223$, $P = 0.0157$); **d**, residual neurocranium outlever length

($F_{2,246} = 13.36$, $P = 3.10 \times 10^{-6}$); **e**, residual buccal cavity length ($F_{2,246} = 12.26$, $P = 8.42 \times 10^{-6}$); **f**, residual gape ($F_{2,246} = 7.974$, $P = 4.41 \times 10^{-4}$). Numbers in parentheses are values of n . Traits are illustrated in Fig. 2e–g. The data conformed reasonably well to parametric statistical assumptions; ANOVA was therefore used to test trait variation among categories.



Extended Data Figure 6 | Relationships between F_2 hybrid functional morphology and niche score. For key functional morphological traits known to differ between wild Paxton benthics and limnetics, trait data from all available F_2 hybrids are plotted against niche score and fitted with linear regression (raw data for gill raker counts; size-corrected data for other traits): **a**, number of long gill rakers ($R^2 = 0.0146$; $F_{1,629} = 9.32$; $P = 0.00236$); **b**, number of short gill rakers ($R^2 = 0.0253$; $F_{1,629} = 16.30$; $P = 6.06 \times 10^{-5}$); **c**, residual anterior epaxial muscle height ($R^2 = 0.0125$; $F_{1,552} = 7.00$;

$P = 0.00804$); **d**, residual anterior epaxial muscle width ($R^2 = 0.0189$; $F_{1,552} = 10.61$; $P = 0.00119$); **e**, residual upper jaw protrusion length ($R^2 = 0.0580$; $F_{1,552} = 34.00$; $P = 9.40 \times 10^{-9}$); **f**, residual lower jaw-opening inlever length ($R^2 = 0.0660$; $F_{1,615} = 43.43$; $P = 9.45 \times 10^{-11}$). Traits are illustrated in Fig. 2e–g. Directions of benthic–limnetic divergence in Paxton Lake (arrows at left of plots, here and in Fig. 2a–d) are based on previously published studies^{16,18,19}, combined with validating counts of long and short gill rakers for this study (data not shown).

Extended Data Table 1 | Goodness-of-fit tests for genomic distribution of QTLs

LG	Size (b.p.)	Percentage of assembled genome's size	Coding + non-coding genes	Percentage of total count of coding + non-coding genes	QTL for all traits	Percentage of total QTL count (all traits)	QTL for component traits only	Percentage of total QTL count (component traits)
1	28,185,914	7.033 %	1,257	6.581 %	2	2.632 %	2	4.878 %
2	23,295,652	5.812 %	860	4.502 %	7	9.210 %	2	4.878 %
3	16,798,506	4.191 %	932	4.880 %	0	0 %	0	0 %
4	32,632,948	8.142 %	1,323	6.926 %	13*,†	17.105 %	5	12.195 %
5	12,251,397	3.057 %	732	3.832 %	0	0 %	0	0 %
6	17,083,675	4.263 %	721	3.775 %	1	1.316 %	0	0 %
7	27,937,443	6.970 %	1,320	6.911 %	7	9.210 %	3	7.317 %
8	19,368,704	4.833 %	881	4.612 %	5	6.579 %	4	9.756 %
9	20,249,479	5.052 %	1,012	5.298 %	4	5.263 %	2	4.878 %
10	15,657,440	3.907 %	815	4.267 %	1	1.316 %	1	2.439 %
11	16,706,052	4.168 %	1,058	5.539 %	1	1.316 %	0	0 %
12	18,401,067	4.591 %	1,003	5.251 %	5	6.579 %	4	9.756 %
13	20,083,130	5.011 %	970	5.078 %	2	2.632 %	1	2.439 %
14	15,246,461	3.804 %	736	3.853 %	5	6.579 %	2	4.878 %
15	16,198,764	4.042 %	778	4.073 %	0	0 %	0	0 %
16	18,115,788	4.520 %	801	4.194 %	9*,†	11.842 %	6*,†	14.634 %
17	14,603,141	3.644 %	702	3.675 %	3	3.947 %	1	2.439 %
18	16,282,716	4.063 %	762	3.989 %	1	1.316 %	0	0 %
19	20,240,660	5.050 %	1,044	5.466 %	0*,†	0 %	0	0 %
20	19,732,071	4.923 %	931	4.874 %	6	7.895 %	5*,†	12.195 %
21	11,717,487	2.924 %	463	2.424 %	4	5.263 %	3†	7.317 %
Sum	400,788,495‡	100 %	19,101§	100 %	76	100 %	41	100 %

Expected numbers of QTLs on LGs, under a random-distribution null hypothesis (simple proportional model), were based on the known size (second column from the left) and gene content (fourth column; predicted number of coding plus non-coding genes) of corresponding chromosomes (obtained from Ensembl genome browser on 17 July 2013; based on initial *G. aculeatus* assembly, Broad S1, February 2006). Observed numbers and percentages of QTLs for all measured traits or only component traits are given in the last four columns (at the right). Results of all tests support the alternative hypothesis of QTL clustering: $\chi^2_{20} = 45.17$, $P = 0.0016$ (for all traits, with a null expectation based on chromosome size); $\chi^2_{20} = 55.76$, $P = 0.0002$ (all traits, based on gene number); $\chi^2_{20} = 34.87$, $P = 0.0219$ (component traits, based on chromosome size); $\chi^2_{20} = 39.12$, $P = 0.0083$ (component traits, based on gene number); P values were estimated by Monte Carlo simulation (10,000 replicates each) because of small expected counts for many LGs. Standardized residuals were used to identify LGs with QTL counts deviating from random expectation (Supplementary Discussion): * $P < 0.05$ (expectation based on chromosome size); † $P < 0.05$ (gene content). Sums for size (‡) and gene content (§) exclude unassembled regions of the genome.

Extended Data Table 2 | MQM model of only main effects of 14 candidate morphological QTLs on niche score

Model term	Map location (cM)	Nearest SNP	PVE	<i>F</i>	d.f.	<i>P</i> -value (<i>F</i>)
LG1	36.00	chrI:25560380	0.52 %	1.688	2	0.185991
LG2†	12.00	chrII:10092618	1.69 %	5.470	2	0.004480†
LG4‡	28.76	chrIV:10997988	2.34 %	7.542	2	0.000596‡
LG7‡	26.99	chrVII:19857837	3.28 %	10.58	2	3.21x10 ⁻⁰⁵ ‡
LG8‡	18.00	chrVIII:16299555	3.74 %	12.08	2	7.63x10 ⁻⁰⁶ ‡
LG9	28.38	chrIX:6126845	0.84 %	2.719	2	0.066968
LG10	4.00	chrX:1275840	0.66 %	2.121	2	0.121029
LG12	38.00	chrXII:15046849	0.57 %	1.846	2	0.159010
LG13*	20.00	chrXIII:17392141	1.16 %	3.733	2	0.024619*
LG14	39.55	chrXIV:4632223	0.24 %	0.776	2	0.460631
LG16†	13.52	chrXVI:9981125	2.16 %	6.989	2	0.001020†
LG17	12.00	chrXVII:2232080	0.03 %	0.109	2	0.896894
LG20*	14.62	chrXX:9279241	1.43 %	4.630	2	0.010198*
LG21	18.00	chrXXI:11060209	0.33 %	1.080	2	0.340247

The table summarizes a main effects-only MQM model enforced to contain all the selected candidate morphological QTLs for niche score (run in R/qtl: niche score as response variable, Haley–Knott regression, with F_2 family covariate). Model terms (at the left) are named according to LG locations of the candidate QTLs, which were limited to the one best candidate for each LG before modelling (see Methods). For each QTL (model term), the table also gives the map position in centimorgans (cM), the nearest SNP marker, the percentage of total variance explained (PVE) for niche score, the *F*-test statistic, the corresponding degrees of freedom (d.f.), and the *P* value. Significant model terms are indicated as follows: * $0.01 \leq P < 0.05$; † $0.001 \leq P < 0.01$; ‡ $P < 0.001$. Overall model results (SS, sum of squares): $SS_{\text{model}} = 169.34$; $d.f._{\text{model}} = 56$; $SS_{\text{error}} = 464.09$; $d.f._{\text{error}} = 473$; $LOD_{\text{model}} = 35.80$; $PVE_{\text{model}} = 26.73\%$; $P\text{value}(F) = 2.91 \times 10^{-11}$.

Extended Data Table 3 | Full MQM model of main QTL effects and effects of pairwise QTL interactions on niche score

Model term	PVE	<i>F</i>	d.f.	<i>P</i> -value (<i>F</i>)
LG1†	3.02 %	3.486	6	0.00222†
LG2†	1.50 %	5.200	2	0.00584†
LG4‡	4.30 %	4.958	6	6.21x10 ⁻⁰⁵ ‡
LG7‡	4.63 %	5.341	6	2.40x10 ⁻⁰⁵ ‡
LG8‡	3.48 %	12.04	2	7.97x10 ⁻⁰⁶ ‡
LG9*	0.96 %	3.328	2	0.03672*
LG13*	1.18 %	4.099	2	0.01720*
LG14*	1.90 %	2.192	6	0.04268*
LG16‡	5.57 %	3.858	10	4.87x10 ⁻⁰⁵ ‡
LG17†	3.95 %	2.737	10	0.00278†
LG20†	1.43 %	4.956	2	0.00742†
LG1xLG16†	2.48 %	4.284	4	0.00205†
LG4xLG17†	1.99 %	3.440	4	0.00873†
LG7xLG17*	1.54 %	2.658	4	0.03233*
LG14xLG16*	1.36 %	2.414	4	0.04817*

The table summarizes the final MQM model of candidate QTL effects on niche score, obtained by the stepwise model selection procedure described in Methods. At all steps, model fitting was performed in R/qtl (niche score as response variable, Haley–Knott regression, with F_2 family covariate). Model terms (at the left) are named according to the LG locations of candidate morphological QTLs (map positions in Extended Data Table 2). For each term (QTL), the table also gives the PVE for niche score, the *F*-test statistic, the corresponding degrees of freedom (d.f.), and the *P*-value. Significant model terms are indicated as follows: * $0.01 \leq P < 0.05$; † $0.001 \leq P < 0.01$; ‡ $P < 0.001$. Overall model results (SS, sum of squares): $SS_{\text{model}} = 209.75$; $d.f._{\text{model}} = 66$; $SS_{\text{error}} = 423.68$; $d.f._{\text{error}} = 463$; $LOD_{\text{model}} = 46.28$; $PVE_{\text{model}} = 33.11\%$; $P \text{ value } (F) = 3.44 \times 10^{-15}$.

Reprogramming human endothelial cells to haematopoietic cells requires vascular induction

Vladislav M. Sandler¹, Raphael Lis^{1,2}, Ying Liu¹, Alon Kedem^{1,2}, Daylon James^{1,2}, Olivier Elemento³, Jason M. Butler¹, Joseph M. Scandura⁴ & Shahin Rafii¹

Generating engraftable human haematopoietic cells from autologous tissues is a potential route to new therapies for blood diseases. However, directed differentiation of pluripotent stem cells yields haematopoietic cells that engraft poorly. Here, we have devised a method to phenocopy the vascular-niche microenvironment of haemogenic cells, thereby enabling reprogramming of human endothelial cells into engraftable haematopoietic cells without transition through a pluripotent intermediate. Highly purified non-haemogenic human umbilical vein endothelial cells or adult dermal microvascular endothelial cells were transduced with the transcription factors *FOSB*, *GFI1*, *RUNX1* and *SPI1* (hereafter referred to as FGRS), and then propagated on serum-free instructive vascular niche monolayers to induce outgrowth of haematopoietic colonies containing cells with functional and immunophenotypic features of multipotent progenitor cells (MPPs). These endothelial cells that have been reprogrammed into human MPPs (rEC-hMPPs) acquire colony-forming-cell potential and durably engraft into immune-deficient mice after primary and secondary transplantation, producing long-term rEC-hMPP-derived myeloid (granulocytic/monocytic, erythroid, megakaryocytic) and lymphoid (natural killer and B cell) progenies. Conditional expression of FGRS transgenes, combined with vascular induction, activates endogenous FGRS genes, endowing rEC-hMPPs with a transcriptional and functional profile similar to that of self-renewing MPPs. Our approach underscores the role of inductive cues from the vascular niche in coordinating and sustaining haematopoietic specification and may prove useful for engineering autologous haematopoietic grafts to treat inherited and acquired blood disorders.

Manufacture of autologous, engraftable haematopoietic stem and progenitor cells (HSPCs) offers tremendous therapeutic potential. Using *in vitro* cultures, human pluripotent stem cells can be differentiated into haematopoietic progenitors, which often have limited expansion potential and do not engraft myeloablated recipients^{1–3}. Enforced expression of transcription factors has also been used to reprogram somatic cells into haematopoietic lineages^{4–6}. Employing cellular fusion, we have shown that direct conversion of somatic cells into fetal HSPCs is also feasible⁷. However, these previous efforts have been unable to produce human haematopoietic cells capable of long-term multilineage engraftment^{4–7}. We hypothesized that in addition to transcription factor expression, haematopoietic specification and long-term engraftment may require inductive signals from the microenvironment. Indeed, the central instructive role of tissue-specific endothelial cells⁸ in supporting organ regeneration^{9,10}, including haematopoietic stem-cell (HSC) self-renewal and reconstitution of multilineage haematopoiesis, has recently come to light^{11–18}.

In mammals, definitive HSCs originate in the vascular microenvironment of the aorta–gonad–mesonephros (AGM)^{19–24}, placenta²⁵ and arterial vessels²⁶. Putative HSCs bud off from haemogenic vascular cells lining the dorsal aorta floor and umbilical arteries, where they are in cellular contact with non-haemogenic endothelial cells²⁷. This ontological endothelial-to-haematopoietic transition (EHT) is mediated in part through expression of the transcription factor *RUNX1* (ref. 21), its non-DNA-binding partner core binding factor- β (ref. 28), *GFI1* and *GFI1b* (refs 29, 30). However, the contribution of microenvironmental inductive signals

provided by anatomically distinct niches and tissue-specific vascular niches⁸ within the AGM, fetal liver and placenta remain poorly defined.

We have identified a minimal set of four transcription factors—*FOSB*, *GFI1*, *RUNX1* and *SPI1* (FGRS)—that reprogram full-term human umbilical vein endothelial cells (HUVECs) and human adult dermal microvascular endothelial cells (hDMECs) into haematopoietic cells with long-term MPP activity (rEC-hMPP). The reprogramming was successful only when a unique serum-free vascular niche platform was used. Subsets of rEC-hMPPs were immunophenotypically marked as HSCs and were capable of long-term primary and secondary multilineage engraftment in immunodeficient mice. We demonstrate that constitutive or transient expression of FGRS transcription factors combined with inductive signals from specialized vascular niche cells^{11,31} are essential for efficient conversion of endothelial cells into rEC-hMPPs.

FGRS and vascular induction reprogramming

Primitive HSCs emerge on a vascular bed during development. Thus, we hypothesized that executive functions of the vascular niche could have an important role during reprogramming by inducing and maintaining nascent haematopoietic cells. Since serum impairs vascular function and interferes with expansion of HSCs and MPPs, we devised a vascular niche model in which endothelial cells transduced with the adenoviral *E4ORF1* gene (E4ECs, VeraVecs) could be cultured without serum^{1,11,12,31}. *E4ORF1* activates survival pathways in endothelial cells without provoking proliferation or cellular transformation and thereby maintains tissue-specific

¹Ansary Stem Cell Institute, Department of Genetic Medicine, and Howard Hughes Medical Institute, Weill Cornell Medical College, New York, New York 10065, USA. ²Ronald O. Perelman and Claudia Cohen Center for Reproductive Medicine, Weill Cornell Medical College, New York, New York 10065, USA. ³HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Cornell Medical College, New York, New York 10065, USA. ⁴Department of Medicine, Hematology-Oncology, Weill Cornell Medical College and the New York Presbyterian Hospital, New York, New York 10065, USA.

functional and metabolic attributes of endothelial cells. E4ECs derived from HUVECs^{1,11,12,31} or endothelial cells purified and propagated from haematopoietic organs^{32,33} balance self-renewal and differentiation of human and mouse long-term HSCs and MPPs by producing physiological levels of Notch ligands, Kit ligand, BMPs, Wnts and other angiocrine factors¹⁴.

To identify transcription factors that drive EHT, we first identified transcription factors differentially expressed by Lin[−]CD34⁺ umbilical cord HSPCs, but not by HUVECs (Extended Data Fig. 1a–d). We then cultivated CD45[−]CD133[−]c-Kit[−]CD31⁺ HUVECs that were devoid of haemogenic potential³⁴ (Fig. 1a) and transduced them with lentiviral vectors expressing various combinations of differentially expressed transcription factor transcripts using GFP as a marker. After 3 days, transduced HUVECs were re-plated onto subconfluent serum-free E4EC monolayers, to force cellular interaction of HUVECs potentially undergoing EHT with inductive vascular niche cells. Within 2 weeks of co-culture with E4ECs, round GFP⁺CD45⁺ cells began to bud from transduced HUVECs and form grape-like colonies (Fig. 1b). Systematic one-by-one dropout of candidate transcription factors demonstrated that expression of FGRS transcription factors was necessary and sufficient for haematopoietic reprogramming of HUVECs (Extended Data Fig. 1b, c).

Co-culture of FGRS-transduced endothelial cells with E4EC monolayers augmented the yield and stability of the haematopoietic-like colonies, which displayed morphological features of haematopoietic progenitors (Fig. 1c). Within 4 weeks of co-culture with E4ECs, FGRS-transduced endothelial cells began to proliferate and form GFP⁺CD45⁺ colonies (Fig. 1a, c). Serum suppressed colony formation and naive HUVECs could not survive without serum and failed to support the emergence of CD45⁺ cells (Fig. 1d). FGRS transduction of 5×10^4 HUVECs followed by 3 weeks of serum-free co-culture with E4ECs yielded 32.3 ± 10.5 colonies (Fig. 1d) (efficiency of reprogramming is 1.5%; see Methods), occasionally forming multi-colony structures (Extended Data Fig. 2a). Once colonies formed, proliferation of GFP⁺ cells increased and after 5 weeks of co-culture with E4ECs, up to 20×10^6 GFP⁺CD45⁺ cells were produced, a ~400-fold expansion of the input FGRS-transduced endothelial cells (Fig. 1d). Since clonal CD45⁺ cells, but not CD45[−] cells, form colonies it is unlikely that E4ECs are mistakenly identified as haematopoietic cells (Extended Data Fig. 2b, c). Thus, FGRS-transduced endothelial cells required sustained inductive and supportive signals from the E4EC vascular niche for efficient haematopoietic reprogramming.

Current efforts to differentiate pluripotent stem cells into repopulating haematopoietic cells have had limited success^{1–3}. We hypothesized

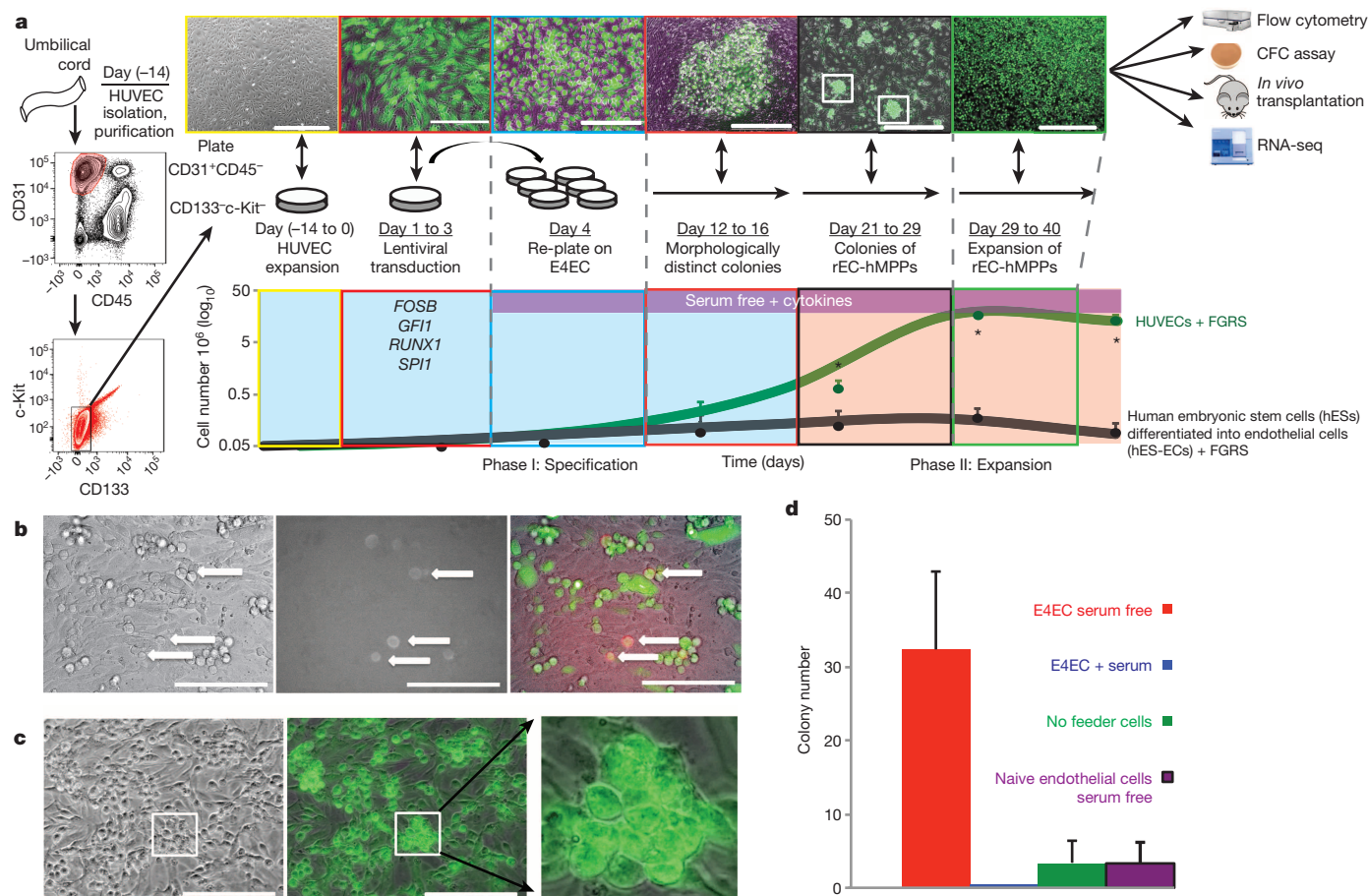


Figure 1 | Reprogramming of HUVECs and hES-ECs into haematopoietic cells by FGRS transcription-factor transduction and vascular induction. **a**, Schema of reprogramming platform of HUVECs into haematopoietic cells. CD45[−]CD31⁺CD133[−]c-Kit[−] cells were sorted from freshly purified HUVECs and expanded (days −14 to 0). Sorted cells were transduced with FGRS (GFP marked) (days 1–3) and grown in endothelial cell media. On day 4, transduced cells were re-plated on E4ECs in serum-free haematopoietic media (days 12–40). Distinct GFP⁺ flat colonies were observed at days 12–16, which by days 21–29 remodelled into three-dimensional grape-like colonies. After a month (days 29–40) GFP⁺ cells expanded ~400-fold ($n = 4$). CD144⁺VEGFR2⁺ endothelial cells derived from hES-ECs³⁵ were also transduced with FGRS. The process of reprogramming is subdivided into two

phases: phase I, specification (day 1–20); phase II, expansion (day 21–40). The expanding cultures were assayed for morphological change, cell number and CD45. Kinetics of reprogramming of HUVECs (green trace) and expansion of reprogrammed hES-ECs cells (black trace) are shown. **b**, Emergence of rounded haematopoietic-like GFP⁺CD45⁺ cells 2–3 weeks after HUVECs were transduced with FGRS (white arrows). **c**, Formation of GFP⁺ haematopoietic-like colonies on the E4ECs 3–4 weeks after FGRS transduction. **d**, Generation of GFP⁺CD45⁺ haematopoietic-like colonies (c) from FGRS-transduced endothelial cells is enhanced by co-culturing with serum-free E4ECs and blocked by the presence of serum ($n = 8$, $P < 0.05$). Scale bar, 200 μ m. Error bars are average \pm s.d.

that FGRS transcription factors could augment haematopoietic differentiation of human embryonic stem (ES) cells. To test this, we first differentiated human ES cells into endothelial cells (hES-EC)^{35,36} and then transduced purified VEGFR2⁺CD144⁺ hES-ECs with FGRS. Although this approach generated CD45⁺CD144⁺ progeny (Extended Data Fig. 2d), these cells did not form stable haematopoietic-like colonies and did not proliferate (Fig. 1a, black line). Thus, hES-ECs are not as permissive as HUVECs for reprogramming into haematopoietic cells.

rEC-hMPPs have features of multilineage progenitors

During reprogramming, GFP⁺ FGRS-transduced endothelial cells and vascular-induced haematopoietic-like colonies lost CD31 expression but gained the expression of human haematopoietic markers hCD45, hCD43, hCD90 (also called Thy-1) and hCD14 (Fig. 2a and Extended Data Fig. 2e). In contrast, the GFP⁺ E4ECs remained CD31⁺CD34⁺CD45⁺. Importantly, a subset of GFP⁺hCD45⁺ FGRS-transduced endothelial cell progeny manifested the immunophenotype of human HSCs (hCD45⁺Lin[−]hCD45RA[−]hCD38[−]hCD90⁺hCD34⁺) and MPPs (hCD45⁺Lin[−]hCD45RA[−]hCD38[−]hCD90[−]hCD34⁺)^{37,38} (Fig. 2b). To assess the function of various populations of these endothelial cells reprogrammed into human MPPs (rEC-hMPPs), we sorted 4-week-old GFP⁺hCD45⁺hCD34⁺ rEC-hMPPs and seeded them in colony-forming cell (CFC)

assays to enumerate progenitor cells. The rEC-hMPPs gave rise to GFP⁺ colonies with CFC-GEMM (granulocytic/erythroid/megakaryocytic/monocytic), CFC-GM (granulocytic/macrophage) and haemoglobinized burst-forming unit-erythroid BFU-E morphologies (Fig. 2c). Flow cytometry and cytospin analysis documented the presence of cells with morphological (Fig. 2d) and immunophenotypic features of granulocyte/macrophage (CD11b⁺, CD14⁺), erythroid (CD235⁺), dendritic (CD83⁺) and megakaryocyte (CD41a⁺) progenies (Extended Data Fig. 2f). The function of rEC-hMPP-derived macrophages was corroborated using a phagocytosis assay (Extended Data Fig. 2g). Thus, rEC-hMPPs contain functional multilineage progenitor cells.

rEC-hMPPs engraft long-term into primary recipients

To assess the engraftment potential of rEC-hMPPs, we transplanted 1.5×10^6 GFP⁺hCD45⁺ rEC-hMPPs into adult sublethally irradiated (275 rad) immunocompromised NSG mice (Fig. 3a). We detected circulating human CD45⁺ cells in the peripheral blood of recipient engrafted mice from 2 to 44 weeks (Fig. 3b) and found hCD45⁺hCD235⁺ erythroid cells 16 weeks post-transplantation (Fig. 3c). We then sorted human CD45⁺ (hCD45⁺) cells from bone marrow of recipient mice 22–24 weeks after transplantation and cultured them for 24 h. The hCD45⁺hCD34⁺ cells were resorted and seeded into CFC assays. They formed CFC-GM,

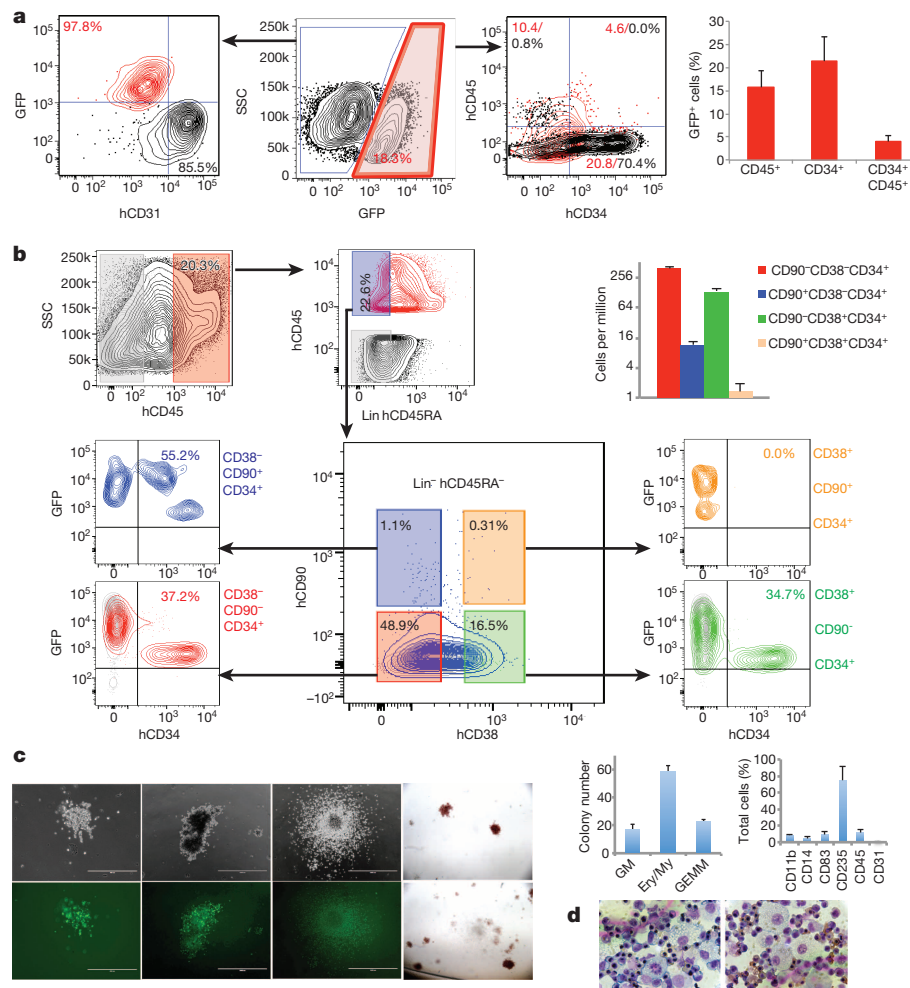


Figure 2 | rEC-hMPPs phenotypically and functionally resemble multilineage HSPCs. **a**, FACS analysis of co-cultured GFP⁺ E4EC vascular niche along with GFP⁺ FGRS-transduced HUVECs 4 weeks after transduction ($n = 9$). **b**, Immunophenotypic analysis of reprogrammed HUVECs (red and blue; $n = 3$). **c**, Four weeks after FGRS transduction and E4EC induction, human GFP⁺hCD45⁺hCD34⁺ cells were sorted and seeded for CFC assay ($n = 3$). Haematopoietic colonies arose in the CFC assay (original

magnification, $\times 4$); wide field (upper row) and corresponding fluorescent images (bottom row) are shown. Left to right: granulocytic-erythroid-monocytic-megakaryocytic (GEMM; scale bar, 400 μ m), erythroid/myeloid, and granulocytic-macrophage (GM) colonies (scale bar, 1,000 μ m), and haemoglobinized colonies (original magnification $\times 4$). Graph shows CFC assay quantification. **d**, Wright-Giemsa stain of cells obtained from the CFC assay colonies. Original magnification, $\times 60$. Error bars are average \pm s.d.

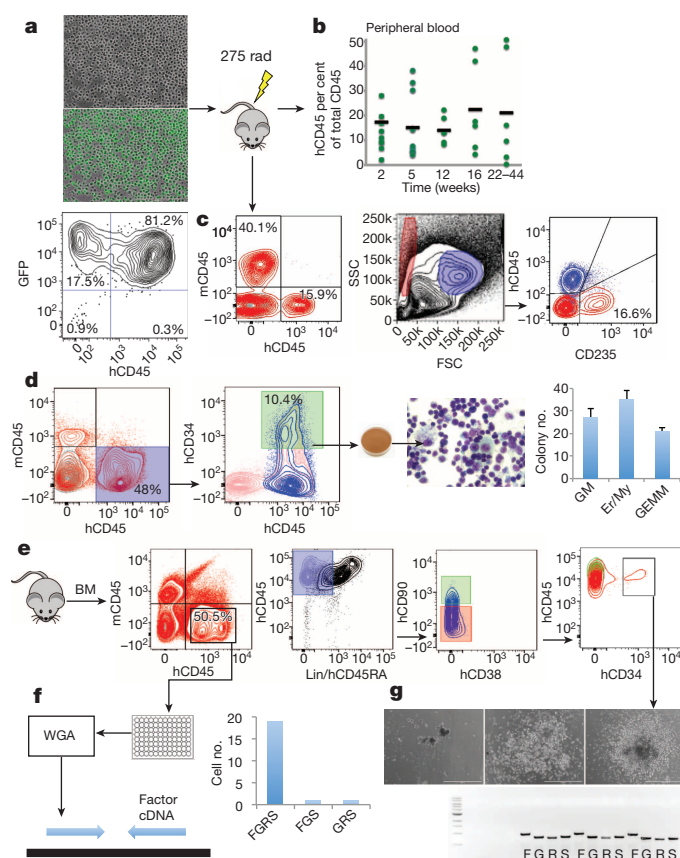


Figure 3 | rEC-hMPPs are capable of *in vivo* erythro-myeloid-megakaryocytic multilineage engraftment. **a**, HUVEC-derived rEC-MPPs were transplanted into sublethally irradiated (275 rad) mice ($n = 9$). **b**, Circulating human CD45⁺ (hCD45⁺) cells were detected at 2 weeks ($n = 7$; $17.38 \pm 7.73\%$), 5 weeks ($n = 6$; $15.1 \pm 13.39\%$), 12 weeks ($n = 6$; $14.14 \pm 5.44\%$), 16 weeks ($n = 6$; $22.36 \pm 17.95\%$) and 22–44 weeks ($n = 6$; $21.23 \pm 22.27\%$) after transplantation. The 22–44 weeks engrafted mice were used for further analyses of the myelodysplasia and fibrotic changes (Extended Data Figs 8, 9 and 10a). **c**, Analysis of the total mononuclear peripheral blood cells at 16 weeks after transplantation of hCD45⁺ and mouse CD45⁺ (mCD45⁺) cells revealed the presence of hCD45⁺ (15.9%) and human non-erythroid circulating cells (left panel). We gated on the FSC/SSC hCD45⁺ erythroid compartment (red gate) and typical human non-erythroid hCD45⁺ compartment (blue gate) (middle and right panels). **d**, rEC-hMPPs isolated from the host retained their multilineage potential *in vitro*; secondary CFC assay. Engraftment of mouse bone marrow 22 weeks after transplantation is shown (left panel). The cells were expanded *in vitro* for 24 h (second panel from left) and FACS resorted for hCD45⁺hCD34⁺ cells for CFC assay. Wright–Giemsa stain (third panel from left) of the cytospin of the cells from CFC assay is shown (original magnification, $\times 100$). The far-right panel shows quantification of the CFCs. Error bars are average \pm s.d. **e**, Phenotypic analysis of *in vivo* 22-weeks engrafted rEC-hMPPs in bone marrow shows human CD45⁺Lin[−]CD45RA[−]CD38[−]CD90[−]CD34⁺ MPPs. **f**, Identification of viral integration on a single-cell level. Whole-genome amplification (WGA) of 21 hCD45⁺ cells isolated from a host mouse (e). Quantification of the analysis is shown in the right graph. **g**, Identification of viral integration on a single-colony level. Lin[−]CD45RA[−]CD38[−]CD90[−]CD34⁺ cells (10 cells) were used for a CFC assay. We detected all four FGRS viral vectors in all CFC colonies tested (bottom image; letters F (*FOSB*), G (*GFI1*), R (*RUNX1*) and S (*SPI1*) show PCR products specific for each of these factors in the right-hand colony; scale bar, 1,000 μ m).

CFC-GEMM and BFU-E haematopoietic colonies with typical myeloid progeny morphologies (Fig. 3d). Hence, rEC-hMPPs are capable of robust multilineage engraftment.

rEC-hMPPs arise from non-haemogenic endothelium

To rule out the possibility that rEC-hMPPs are derived from rare contaminating haemogenic cells^{25,34}, we tested whether naive endothelial cells

or highly purified, mature CD45[−]CD144⁺CD31⁺CD62E⁺ (E-selectin)⁺ endothelial cells^{32,33} could form haematopoietic cells when cultured in optimal pro-haematopoietic media. Neither serum withdrawal nor addition of haematopoietic cytokines induced formation of CD45⁺CD34⁺ cells from naive HUVECs (Extended Data Fig. 3a, b). However, FGRS transduction and E4EC induction of the clonal or oligo-clonal CD45[−]CD144⁺CD31⁺CD62E⁺ mature endothelial cells^{32,33} generated functional rEC-MPPs with high efficiency (Extended Data Figs 3c–e and 4a–c). Thus, rEC-hMPPs are not derived from a scarce population of spontaneously differentiating endothelial cells with pre-existing haemogenic potential.

The bone marrow of robustly engrafted recipient NSG mice contained a small population of hCD45⁺ cells with Lin[−]CD45RA[−]CD38[−]CD90[−]CD34⁺ immunophenotype of human MPPs³⁷ (Fig. 3e). To ensure that engrafted cells were derived from FGRS-transduced endothelial cells, we purified hCD45⁺ cells from recipient bone marrow (Fig. 3f) and seeded single cells into 96-well plates for whole-genome amplification (WGA) and detection of viral vector integration. All hCD45⁺ cells had been transduced by lentiviral vectors, and 19 of 21 cells showed integration of all four FGRS transcription factors (Fig. 3f and Extended Data Fig. 5). To verify that these cells were the progeny of rEC-hMPPs, we seeded hCD45⁺ cells for CFC assays to examine viral integration in individual colonies (Fig. 3g). We demonstrated that all tested colonies were derived from cells that had integrated the lentiviral vectors expressing FGRS (Fig. 3g). Therefore, engrafted human haematopoietic cells were derived from transplanted rEC-hMPPs.

To test whether FGRS-induced reprogramming triggered expression of endogenous human FGRS genes²⁴, we expressed genetically distinct murine transcription factors (mFGRS) using inducible lentiviral vectors to reprogram HUVECs into rEC-hMPPs and then assessed expression of the endogenous human FGRS genes. Transient expression of mFGRS with E4EC co-culture for 3 weeks induced a 100-fold greater expression of endogenous genes than that of switched-off mFGRS transcripts (Extended Data Fig. 6a–c). Therefore, rEC-hMPPs do not require continuous expression of exogenous FGRS transcription factors to sustain their haematopoietic cell fates.

Furthermore, we speculated that enforced SPI1 expression might prevent rEC-hMPPs from differentiating into T cells^{39,40}. Therefore, we constitutively expressed FGR transcription factors with a Tet-inducible *SPI1* and E4EC induction and identified a small but significant population of CD3⁺ T cells (Extended Data Fig. 6d, e). Thus, generation of lymphoid cells from rEC-hMPPs could be optimized by transient expression of specific transcription factors.

Reprogramming adult endothelial cells to rEC-hMPPs

To test whether our approach could reprogram adult human endothelial cells, we transduced human DMECs (hDMECs) with FGRS transcription factors and propagated them on serum-free E4EC monolayers (Fig. 4a). After 4 weeks, GFP⁺hCD45⁺hCD34⁺ cells were sorted for CFC assay. The rEC-hMPPs yielded cells with morphological features of haematopoietic cells (Fig. 4b, left panel) and functional myeloid CFC-GM, CFC-GEMM and BFU-Es (Fig. 4b, middle panel), containing CD235⁺ erythroid, hCD33⁺hCD14⁺hCD11b⁺ macrophage/monocyte, and hCD83⁺ dendritic cell progenies (Fig. 4b, middle panel, and Extended Data Fig. 7a).

Next, we compared the transcriptional profiles of rEC-hMPPs before and after NSG engraftment to those of naive HUVECs, hDMECs and purified Lin[−]CD34⁺ cord blood HSPCs (Fig. 4c). FGRS transduction plus E4EC induction activated haematopoietic genes and downregulated vascular gene signatures (Fig. 4c). Importantly, 22 weeks post-transplantation, CD45⁺CD34⁺ rEC-hMPPs had a transcription pattern similar to normal hCD34⁺ cord blood HSPCs and distinct from the endothelial cells from which they were derived (Fig. 4c). Notably, pluripotency genes were not induced in rEC-hMPPs, indicating that reprogramming does not require transition through a destabilizing pluripotent intermediate (Fig. 4d).

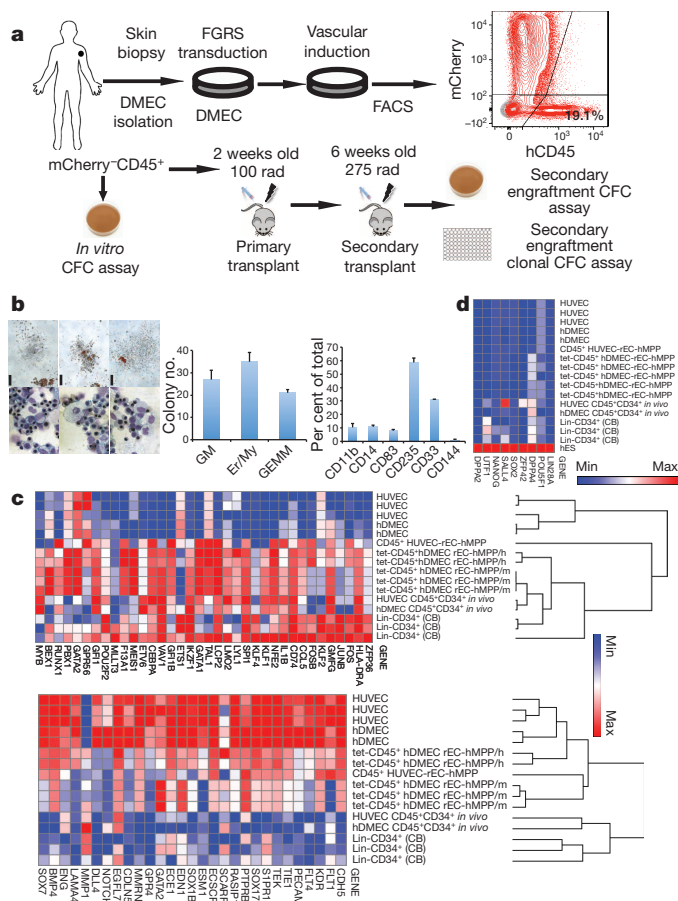


Figure 4 | Functional and transcriptional analysis of adult hDMEC-derived rEC-hMPPs. **a**, Schematic representation of *in vitro* and *in vivo* functional tests of hDMEC-derived rEC-hMPPs. **b**, Left: haematopoietic colonies observed in CFC assay (scale bar, 200 μm); wide field (upper row). Wright–Giemsa stain of cells from CFC colonies (original magnification, ×60) is shown in the bottom row. Middle: quantification of the CFC assay (*n* = 3). Right: immunophenotypic quantification of surface marker expression in CFC colonies (*n* = 3). **c**, Comparison of the global gene transcription profiles of primary HUVECs and hDMECs; rEC-hMPPs reprogrammed from HUVECs after 4 weeks (CD45⁺HUVEC-rEC-hMPPs); rEC-hMPPs reprogrammed from hDMECs by transduction with inducible Tet-On murine FGFR3 (/m) or human FGFR3 (/h) and vascular induction for 3–4 weeks (tet-CD45⁺hDMEC rEC-hMPPs); engrafted human CD45⁺CD34⁺ cells purified from the bone marrow of NSG mice 22 weeks after primary transplantation with HUVEC-reprogrammed rEC-hMPPs (HUVEC CD45⁺CD34⁺ *in vivo*); engrafted human CD45⁺CD34⁺ cells purified from the bone marrow of NSG mice 15 weeks after secondary engraftment with hDMEC-reprogrammed rEC-hMPPs (hDMEC CD45⁺CD34⁺ *in vivo*); and naive purified Lin[−]CD34⁺ cells from cord blood (CB). **d**, Comparison of expression of prototypical pluripotency genes shown in c. hES, human embryonic stem cells. The data in c and d are presented as log₂ (transcription level). Error bars are average ± s.d.

rEC-hMPPs engraft primary and secondary recipients

To assess the engraftment potential of hDMEC-derived rEC-hMPPs, we transplanted 1×10^5 CD45⁺GFP⁺ rEC-hMPPs into sublethally irradiated (100 rad) 2-week-old neonatal NSG mice (Fig. 5a). Circulating hCD45⁺ cells were detected in the peripheral blood of recipient animals 4 weeks ($2.09 \pm 1.27\%$), 6 weeks ($4.46 \pm 3.66\%$) and 12 weeks ($4.05 \pm 3.50\%$) after transplantation (Fig. 5a). Fourteen weeks post-transplantation, human haematopoietic cells were found in peripheral blood, bone marrow and spleen (Fig. 5a and Extended Data Fig. 7b–d). Notably, these recipient animals harboured myeloid and lymphoid populations, including hCD19⁺ B cells ($10.13 \pm 4.98\%$), hCD56⁺ natural killer cells ($1.62 \pm 0.67\%$), hCD11b⁺ monocyte/macrophages ($27.66 \pm 8.92\%$) and hCD41a⁺ megakaryocytes ($4.90 \pm 1.51\%$) in their spleens (Fig. 5a and Extended

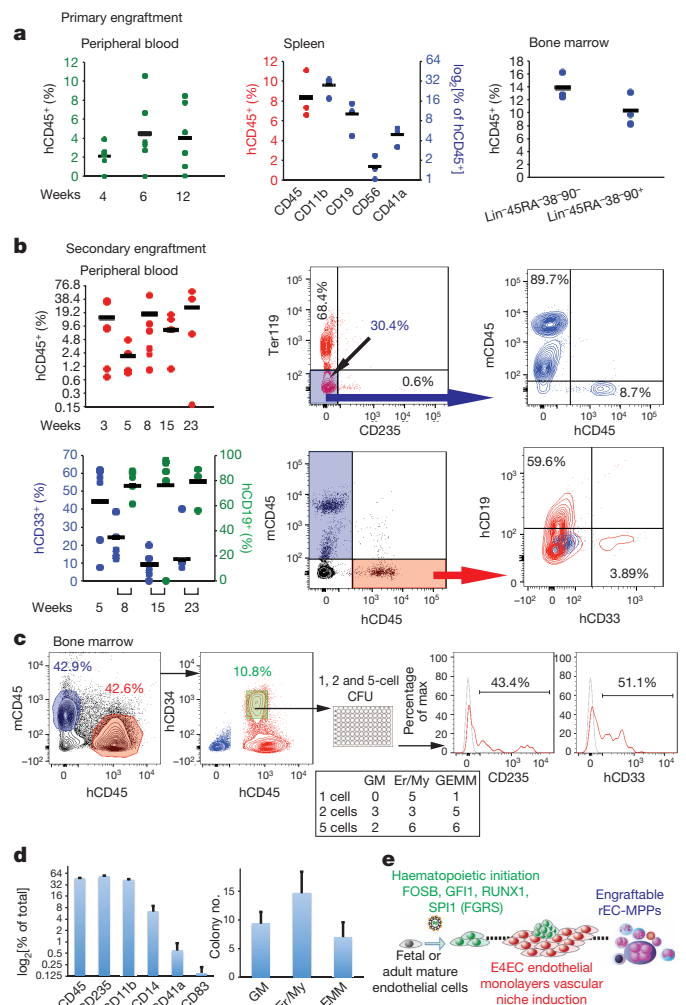


Figure 5 | Adult human hDMEC-derived rEC-hMPPs are capable of *in vivo* primary and secondary multilineage engraftment. **a**, Analysis of peripheral blood of mice at 4, 6 and 12 weeks after primary transplantation (*n* = 6). Analysis of spleen and bone marrow of mice at 14 weeks after primary transplantation (*n* = 3). **b**, Analysis of the peripheral blood of mice at 3 weeks, 5 weeks, 8 weeks (for all time points; *n* = 6), 15 weeks (*n* = 4) and 23 weeks (*n* = 4) after secondary transplantation (*n* = 6). FACS plots on the right show representative analysis of rEC-hMPP secondary engraftment. Mouse Ter119⁺ and human CD235⁺ erythroid populations were excluded to obtain an accurate estimation of hCD45⁺ and mCD45⁺ cells. **c**, Clonal CFC assay of bone marrow hCD45⁺hCD34⁺ cells (*n* = 3; left plot). Emerging colonies were counted and classified (middle table). CFC colonies derived from single plated hCD45⁺hCD34⁺ cells comprise mixed-lineage erythroid and myeloid progenies (right plots). **d**, Reprogrammed cells isolated from the host retained their multilineage potential *in vitro*; secondary CFC assay. **e**, Schematic representation of steps of reprogramming of endothelial cells into rEC-hMPPs by FGFR3 transcription factors and E4EC vascular niche induction.

Data Fig. 7b–d). Hence, rEC-hMPPs are capable of prolonged multilineage haematopoietic engraftment.

The bone marrow of primary recipient mice (weeks 12–14) contained populations with the immunophenotype of human HSCs (hCD45⁺Lin[−]hCD45RA[−]hCD38[−]hCD90⁺hCD34⁺, $10.37 \pm 2.55\%$) and MPPs (hCD45⁺Lin[−]hCD45RA[−]hCD38[−]hCD90[−]hCD34⁺, $13.83 \pm 2.14\%$) (Extended Data Fig. 7d)^{37,38}. Because these populations can self-renew, we tested whether bone marrow cells of mice engrafted by primary rEC-hMPPs (12 weeks post-transplant) could engraft secondary NSG recipient mice. Indeed, the peripheral blood of secondary recipients was engrafted by human myeloid and lymphoid progenies 3 weeks ($14.61 \pm 15.7\%$), 5 weeks ($2.01 \pm 1.5\%$), 8 weeks ($17.78 \pm 16.23\%$), 15 weeks ($7.99 \pm 7.36\%$)

and 23 weeks ($26.3 \pm 25.7\%$) post-transplantation (Fig. 5b). Thus, subpopulations of rEC-hMPPs can self-renew and are capable of durable myeloid and lymphoid engraftment in NSG mice—characteristics similar to true hMPPs.

To examine whether individual rEC-hMPPs retained clonal multilineage potential, we isolated hCD45⁺hCD34⁺ cells from the bone marrow of a secondary robustly engrafted mouse 15 weeks post-transplantation and then assessed the multilineage CFC activity in clonal (1 cell per well), oligo-clonal (2 and 5 cells per well) and bulk (1,000 cells per well) sorted cells (Fig. 5c, d). All single-cell-derived colonies displayed multilineage differentiation, including hCD33⁺hCD14⁺hCD11b⁺ myeloid, hCD41⁺ megakaryocytic and CD235⁺ erythroid progenies (Fig. 5c, d), indicating that engrafted rEC-hMPPs from secondary transplantants retained their MPP potential. Thus, individual cells within the rEC-hMPPs have the immunophenotypic and functional attributes of HSPC-like/self-renewing hMPPs (Fig. 5e).

Notably, rEC-hMPPs isolated from primary and secondary engrafted mice showed no evidence of malignant transformation (Extended Data Figs 8, 9, 10a) or genetic abnormalities (Extended Data Fig. 10b).

Discussion

The availability of engraftable autologous human cells offers the potential to cure a wide spectrum of benign and malignant haematological disorders. Previous efforts using pluripotent stem cells have been handicapped by low efficiency and poor engraftment^{2,3,41,42}. Here, we have taken advantage of an ontological link between endothelial and haematopoietic cells to efficiently reprogram mature, fetal and adult endothelial cells into engraftable self-renewing hMPPs without transitioning through a potentially destabilizing pluripotent intermediate. Just as support from non-haemogenic vascular cells is important for EHT during development, we found that the instructive contribution of the vascular niche was central to reprogramming endothelial cells to haematopoietic cells.

Differentiating pluripotent stem cells or expanding AGM-derived cells to engraftable haematopoietic cells has been inefficient when stromal cells have been used for niche-like support^{2,3,41}. This could be due to: (1) poor inductive function of stromal cells in serum-free culture; and/or (2) distinguishing features of endothelial cells that resemble the haematogenic niche cells that support EHT^{11–15}. For example, E4ECs produce the proper physiological levels of inductive angiocrine factors, including Notch, BMP and c-Kit pathways¹⁴ that are important for EHT⁴³. Thus, adult organ-specific pro-haematopoietic vascular niches, such as HUVECs^{11–18}, bone marrow^{11–18,32,33}, hepatic and splenic sinusoids⁸ may share functional characteristics with EHT-inductive niche cells.

The rEC-hMPPs can engraft primary and secondary recipient mice with individual cells capable of differentiating to multiple haematopoietic lineages. However, the recipient microenvironmental signals and temporal aspects of reprogramming influence the outcome of xenograft studies. NSG mice lack the proper niches for T-cell differentiation and we were not able to determine whether engrafted rEC-hMPPs could give rise to T cells *in vivo*. We found that temporally restricted expression of *SPI1*, along with sustained FGR, increased lymphoid differentiation of the rEC-hMPPs, suggesting that sustained *SPI1* interfered with lymphogenesis. Notably, even transient expression of FGRS is sufficient to activate endogenous transcription factors. The age of recipient mice was also important because transplantation of neonatal (2-week-old) NSG mice enhanced lymphoid engraftment by rEC-MPPs. Therefore, temporal and chronological expression of FGRS transcription factors with proper stoichiometry combined with vascular niche signals appears to increase the yield of rEC-hMPPs with authentic multilineage, long-term, self-renewing HSPC function.

Direct reprogramming of endothelial cells into engraftable HSPCs coordinated by the inductive signals conveyed by tissue-specific vascular niches offers an innovative way to decipher the hierarchy of transcription factors and microenvironmental cues that guide haematopoietic development. Our approach lays a foundation for engineering engraftable

autologous rEC-hMPPs and potentially true HSCs for treatment of patients with haematological disorders.

METHODS SUMMARY

Endothelial cells were reprogrammed into haematopoietic cells by transduction with transcription factors and vascular niche induction. To establish vascular niche platform, endothelial cells were purified and transduced with a lentiviral vector expressing the adenoviral *E4ORF1* gene (E4ECs, VeraVecs, Angiocrine Bioscience, New York, NY). Purified CD45⁺CD133⁺c-Kit⁺CD31⁺ and clonal populations of CD45⁺CD144⁺CD31⁺CD62E⁺ full-term human umbilical vein endothelial cells (HUVECs) and adult primary human dermal microvascular endothelial cells (hDMEC) were cultured in endothelial cell growth medium. Then, HUVECs or hDMECs were transduced with lentiviral vectors expressing GFP and a combination of transcription factors: *FOSB*, *GFI1*, *RUNX1* and *SPI1* (FGRS). After 3 days, GFP⁺ FGRS-transduced endothelial cells were plated in co-culture with 30–50% subconfluent E4EC monolayers supplemented with serum-free haematopoietic media composed of StemSpan SFEM, 10% KnockOut serum replacement, 5 ng ml^{−1} FGF-2, 10 ng ml^{−1} EGF, 20 ng ml^{−1} SCF, 20 ng ml^{−1} FLT3, 20 ng ml^{−1} TPO, 20 ng ml^{−1} IGF-1, 10 ng ml^{−1} IGF-2, 10 ng ml^{−1} IL-3 and 10 ng ml^{−1} IL-6. After 3–4 weeks of co-culture, outgrown GFP⁺ reprogrammed endothelial cells into human multipotent progenitor cells (rEC-hMPPs) formed typical grape-like haematopoietic colonies. After 4 weeks, human CD45⁺ rEC-hMPPs were FACS sorted for: (1) immunophenotypic analyses; (2) methylcellulose-CFC assay; (3) molecular profiling; (4) comparative genomic hybridization; and (5) transplanted retro-orbitally into primary sublethally irradiated (275 rad) 6-week-old NSG mice or sublethally irradiated (100 rad) 2-week-old mice neonates. After 3 months, sorted, bone-marrow-derived human CD45⁺ cells (hCD45⁺ cells) or whole bone marrow of the primary engrafted mice were transplanted into secondary recipients. After 3 months of primary and 6 months of the secondary transplantation, engrafted hCD45⁺ cells in bone marrow, spleen and peripheral blood of mice were FACS sorted and processed for: (1) multivariate immunophenotypic analyses; (2) clonal and oligo-clonal CFC assay; and (3) molecular profiling. Tissues of the engrafted mice were processed for histological examination to rule out malignant transformation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 April 2013; accepted 3 June 2014.

Published online 2 July 2014.

1. Rafii, S. *et al.* Human ESC-derived hemogenic endothelial cells undergo distinct waves of endothelial to hematopoietic transition. *Blood* **121**, 770–780 (2013).
2. Sturgeon, C. M., Ditadi, A., Clarke, R. L. & Keller, G. Defining the path to hematopoietic stem cells. *Nature Biotechnol.* **31**, 416–418 (2013).
3. Choi, K. D. *et al.* Identification of the hemogenic endothelial progenitor and its direct precursor in human pluripotent stem cell differentiation cultures. *Cell Rep.* **2**, 553–567 (2012).
4. Pereira, C. F. *et al.* Induction of a hemogenic program in mouse fibroblasts. *Cell Stem Cell* **13**, 205–218 (2013).
5. Szabo, E. *et al.* Direct conversion of human fibroblasts to multilineage blood progenitors. *Nature* **468**, 521–526 (2010).
6. Xie, H., Ye, M., Feng, R. & Graf, T. Stepwise reprogramming of B cells into macrophages. *Cell* **117**, 663–676 (2004).
7. Sandler, V. M., Lailier, N. & Bouhassira, E. E. Reprogramming of embryonic human fibroblasts into fetal hematopoietic progenitors by fusion with human fetal liver CD34⁺ cells. *PLoS ONE* **6**, e18265 (2011).
8. Nolan, D. J. *et al.* Molecular signatures of tissue-specific microvascular endothelial cell heterogeneity in organ maintenance and regeneration. *Dev. Cell* **26**, 204–219 (2013).
9. Ding, B. S. *et al.* Inductive angiocrine signals from sinusoidal endothelium are required for liver regeneration. *Nature* **468**, 310–315 (2010).
10. Ding, B. S. *et al.* Endothelial-derived angiocrine signals induce and sustain regenerative lung alveolarization. *Cell* **147**, 539–553 (2011).
11. Butler, J. M. *et al.* Endothelial cells are essential for the self-renewal and repopulation of Notch-dependent hematopoietic stem cells. *Cell Stem Cell* **6**, 251–264 (2010).
12. Butler, J. M. *et al.* Development of a vascular niche platform for expansion of repopulating human cord blood stem and progenitor cells. *Blood* **120**, 1344–1347 (2012).
13. Hooper, A. T. *et al.* Engraftment and reconstitution of hematopoiesis is dependent on VEGFR2-mediated regeneration of sinusoidal endothelial cells. *Cell Stem Cell* **4**, 263–274 (2009).
14. Kobayashi, H. *et al.* Angiocrine factors from Akt-activated endothelial cells balance self-renewal and differentiation of haematopoietic stem cells. *Nature Cell Biol.* **12**, 1046–1056 (2010).
15. Poulos, M. G. *et al.* Endothelial jagged-1 is necessary for homeostatic and regenerative hematopoiesis. *Cell Rep.* **4**, 1022–1034 (2013).

16. Avecilla, S. T. *et al.* Chemokine-mediated interaction of hematopoietic progenitors with the bone marrow vascular niche is required for thrombopoiesis. *Nature Med.* **10**, 64–71 (2004).
17. Ding, L., Saunders, T. L., Enikolopov, G. & Morrison, S. J. Endothelial and perivascular cells maintain haematopoietic stem cells. *Nature* **481**, 457–462 (2012).
18. Doan, P. L. *et al.* Tie2⁺ bone marrow endothelial cells regulate hematopoietic stem cell regeneration following radiation injury. *Stem Cells* **31**, 327–337 (2013).
19. Orkin, S. H. & Zon, L. I. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* **132**, 631–644 (2008).
20. Medvinsky, A. & Dzierzak, E. Definitive hematopoiesis is autonomously initiated by the AGM region. *Cell* **86**, 897–906 (1996).
21. North, T. E. *et al.* Runx1 expression marks long-term repopulating hematopoietic stem cells in the midgestation mouse embryo. *Immunity* **16**, 661–672 (2002).
22. Yoshimoto, M., Porayette, P. & Yoder, M. C. Overcoming obstacles in the search for the site of hematopoietic stem cell emergence. *Cell Stem Cell* **3**, 583–586 (2008).
23. Eilken, H. M., Nishikawa, S. & Schroeder, T. Continuous single-cell imaging of blood generation from haemogenic endothelium. *Nature* **457**, 896–900 (2009).
24. Swiers, G. *et al.* Early dynamic fate changes in haemogenic endothelium characterized at the single-cell level. *Nature Commun.* **4**, 2924 (2013).
25. Rhodes, K. E. *et al.* The emergence of hematopoietic stem cells is initiated in the placental vasculature in the absence of circulation. *Cell Stem Cell* **2**, 252–263 (2008).
26. Gordon-Keylock, S., Sobiesiak, M., Rytsov, S., Moore, K. & Medvinsky, A. Mouse extraembryonic arterial vessels harbor precursors capable of maturing into definitive HSCs. *Blood* **122**, 2338–2345 (2013).
27. Zovein, A. C. *et al.* Fate tracing reveals the endothelial origin of hematopoietic stem cells. *Cell Stem Cell* **3**, 625–636 (2008).
28. Chen, M. J. *et al.* Erythroid/myeloid progenitors and hematopoietic stem cells originate from distinct populations of endothelial cells. *Cell Stem Cell* **9**, 541–552 (2011).
29. Lancrin, C. *et al.* Gfi1 and Gfi1B control the loss of endothelial identity of hemogenic endothelium during hematopoietic commitment. *Blood* **120**, 314–322 (2012).
30. Hock, H. *et al.* Gfi-1 restricts proliferation and preserves functional integrity of haematopoietic stem cells. *Nature* **431**, 1002–1007 (2004).
31. Seandel, M. *et al.* Generation of a functional and durable vascular niche by the adenoviral E4ORF1 gene. *Proc. Natl Acad. Sci. USA* **105**, 19288–19293 (2008).
32. Rafii, S. *et al.* Isolation and characterization of human bone marrow microvascular endothelial cells: hematopoietic progenitor cell adhesion. *Blood* **84**, 10–19 (1994).
33. Rafii, S. *et al.* Human bone marrow microvascular endothelial cells support long-term proliferation and differentiation of myeloid and megakaryocytic progenitors. *Blood* **86**, 3353–3363 (1995).
34. Wu, X., Lensch, M. W., Wylie-Sears, J., Daley, G. Q. & Bischoff, J. Hemogenic endothelial progenitor cells isolated from human umbilical cord blood. *Stem Cells* **25**, 2770–2776 (2007).
35. James, D. *et al.* Expansion and maintenance of human embryonic stem cell-derived endothelial cells by TGF β inhibition is Id1 dependent. *Nature Biotechnol.* **28**, 161–166 (2010).
36. Ginsberg, M. *et al.* Efficient direct reprogramming of mature amniotic cells into endothelial cells by ETS factors and TGF β suppression. *Cell* **151**, 559–575 (2012).
37. Chao, M. P., Seita, J. & Weissman, I. L. Establishment of a normal hematopoietic and leukemia stem cell hierarchy. *Cold Spring Harb. Symp. Quant. Biol.* **73**, 439–449 (2008).
38. Notta, F. *et al.* Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment. *Science* **333**, 218–221 (2011).
39. Zou, G. M., Chen, J. J., Yoder, M. C., Wu, W. & Rowley, J. D. Knockdown of Pu.1 by small interfering RNA in CD34⁺ embryoid body cells derived from mouse ES cells turns cell fate determination to pro-B cells. *Proc. Natl Acad. Sci. USA* **102**, 13236–13241 (2005).
40. Pongubala, J. M. *et al.* Transcription factor EBF restricts alternative lineage options and promotes B cell fate commitment independently of Pax5. *Nature Immunol.* **9**, 203–215 (2008).
41. Ledran, M. H. *et al.* Efficient hematopoietic differentiation of human embryonic stem cells on stromal cells derived from hematopoietic niches. *Cell Stem Cell* **3**, 85–98 (2008).
42. Doulatov, S. *et al.* Induction of multipotential hematopoietic progenitors from human pluripotent stem cells via respecification of lineage-restricted precursors. *Cell Stem Cell* **13**, 459–470 (2013).
43. Marcelo, K. L. *et al.* Hemogenic endothelial cell specification requires c-Kit, Notch signaling, and p27-mediated cell-cycle control. *Dev. Cell* **27**, 504–515 (2013).

Acknowledgements V.M.S. is supported by Empire State Stem Cell Board (ESSCB) and New York State Department of Health (NYSDH C026878). S.R. is supported by Ansary Stem Cell Institute (ASCI), HHMI, ESSCB/NYSDH (C024180, C026438, C026878, C028117), NHLBI (R01HL097797, R01HL119872, U01 HL099997), NIDDK (R01DK095039), NCI (U54CA163167), Qatar National Priorities Research Foundation grant NPRP08-663-3-140 and the Qatar Foundation BioMedical Research Program. J.M.S. is supported by grants from NCI (CA159175 and CA163167), NHLBI (HL119872 and HL055748), Starr Foundation and a Leukemia & Lymphoma Society Scholar award. J.M.B. is supported by an American Society of Hematology Scholar Award, NHLBI U01-HL099997 and Angiocrine Bioscience and ASCI. We acknowledge the contribution of J. Z. Xiang and Agnes J. Viale for enabling and executing molecular profiling, and E. Gars for technical support. We appreciate W. Schachterle for recommendations and edits of the manuscript.

Author Contributions V.M.S. and S.R. conceived and designed the project, performed experiments, analysed the data and wrote the manuscript. R.L., Y.L. and J.M.B. performed experiments, interpreted and analysed data. J.M.S. interpreted, analysed data and wrote the manuscript. D.J., O.E. and A.K. performed the experiments and analysed the data. All authors commented on the paper.

Author Information Data have been deposited in the GEO database under accession number GSE57662. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.R. (srafii@med.cornell.edu).

METHODS

Fetal and adult endothelial cells used for reprogramming. Full-term human umbilical vein endothelial cells (HUVECs) were obtained as previously described^{32,33}. Multiple purified populations of CD45⁺ CD133⁺ c-Kit⁺ CD31⁺ HUVECs were isolated from separate umbilical cords ($n = 8$) and were cultured in endothelial growth media (EM): Medium 199 (Thermo Scientific: FB-01), 20% fetal bovine serum (Omega Scientific), 20 $\mu\text{g ml}^{-1}$ endothelial cell supplement (Biomedical Technologies: BT-203), 1 \times Pen/Strep, and 20 units ml^{-1} heparin (Sigma: H3149-100KU). Multiple batches ($n = 3$) of adult primary human dermal microvascular endothelial cells (hDMECs) were purchased from ScienCell Research Laboratories (catalogue 2020). In addition, cultured HUVECs were passaged for 3–5 times and then CD45⁺ CD144⁺ CD31⁺ CD62E⁺ HUVECs were sorted for clonal analyses to rule out contamination with pre-existing haemogenic endothelial cells.

For reprogramming experiments, transduced HUVECs or hDMECs were co-cultured with E4ECs in serum-free haematopoietic media (HM) formulated as StemSpan SFEM (StemCell Technologies), 10% KnockOut Serum Replacement (Invitrogen), 5 ng ml^{-1} bFGF (FGF-2), 10 ng ml^{-1} EGF, 20 ng ml^{-1} SCF (soluble Kit-ligand), 20 ng ml^{-1} FLT3, 20 ng ml^{-1} TPO, 20 ng ml^{-1} IGF-1, 10 ng ml^{-1} IGF-2, 10 ng ml^{-1} IL-3, 10 ng ml^{-1} IL-6 (all from Invitrogen, eBioscience, or Peprotech).

Manufacture of vascular niche platform. To establish the vascular niche monolayers, HUVECs were purified and transduced with lentiviral vector carrying a cassette of adenoviral *E4ORF1* gene (E4ECs) as previously described³¹ or obtained as VeraVecs from Angiocrine Bioscience, New York, NY. E4ECs proliferate in serum-free and xenobiotic-free conditions only supplemented with minimal angiogenic factors. All naive endothelial cells that are non-transduced with E4ORF1 are depleted during passaging in serum-free conditions. Confluent monolayers of E4ECs are contact inhibited, non-transformed and propagate as homogenous monolayers providing an ideal instructive niche for reprogramming and sustaining FGRS-transduced endothelial cells into rEC-hMPPs.

Reprogramming of endothelial cells into MPPs (rEC-hMPPs). Endothelial cells were reprogrammed into haematopoietic cells by transduction with transcription factors and vascular niche induction. Purified populations of CD45⁺ CD133⁺ c-Kit⁺ CD31⁺ and clonal CD45⁺ CD144⁺ CD31⁺ CD62E⁺ full-term HUVECs and adult primary hDMECs were cultured in the endothelial cell growth medium (EM). Then, HUVECs or hDMECs were transduced with lentiviral vectors expressing GFP and a combination of transcription factors—*FOSB*, *GFI1*, *RUNX1* and *SPI1* (FGRS)—and maintained in EM. After 3 days, GFP⁺ FGRS transduced endothelial cells were plated in co-culture with 30% to 50% subconfluent E4EC monolayers supplemented with serum-free haematopoietic media (HM) composed of StemSpan SFEM, 10% KnockOut Serum Replacement, 5 ng ml^{-1} FGF-2, 10 ng ml^{-1} EGF, 20 ng ml^{-1} SCF, 20 ng ml^{-1} FLT3, 20 ng ml^{-1} TPO, 20 ng ml^{-1} IGF-1, 10 ng ml^{-1} IGF-2, 10 ng ml^{-1} IL-3, 10 ng ml^{-1} IL-6. After 3–4 weeks of co-culture the out-grown GFP⁺ reprogrammed endothelial cells into human multipotent progenitor cells (rEC-hMPPs) formed typical grape-like haematopoietic colonies. At the end of 4 weeks, human CD45⁺ rEC-hMPPs were FACS sorted for: (1) immunophenotypic analyses; (2) methylcellulose-CFC assay (five thousand to ten thousand cells per well); (3) molecular profiling; (4) comparative genomic hybridization; and (5) transplanted retro-orbitally into primary sublethally irradiated (275 rad) 6-week-old NSG mice or sublethally irradiated (100 rad) 2-week-old mice neonates. After 3 months, human CD45⁺ cells (hCD45⁺ cells) derived from bone marrow or whole bone marrow of the primary engrafted mice were used for transplantation into secondary recipients. After 3 months of primary and 6 months of the secondary transplantation, engrafted hCD45⁺ cells in bone marrow, spleen and peripheral blood of mice were FACS sorted and processed for: (1) multivariate immunophenotypic analyses; (2) multi-cell and clonal methylcellulose CFC assay; and (3) molecular profiling. Tissues of the engrafted mice were processed for histological examination to rule out malignant transformation.

Increasing efficiency of reprogramming. To increase efficiency of the reprogramming, we developed a strategy to select those subsets of endothelial cells that were most likely transduced with the proper stoichiometry of all four FGRS transcription factors. We initially focused on generating endothelial cells transduced with *GFI1*, *SPI1* and *FOSB* transcription factors because their expression in endothelial cells is negligible (Extended Data Fig. 1). To accomplish this, we transduced 5×10^6 endothelial cells with FGRS lentiviral 'cocktail' marked by puromycin resistance (*SPI1*) or GFP (*FOSB* and *GFI1*). We then applied puromycin selection for 2 days to enrich *SPI1*-expressing cells and sorted them for GFP expression to enrich for *SPI1*⁺ GFP⁺ (*FOSB*/*GFI1*) endothelial cells. Subsequently, we transduced these GFP⁺ puromycin-resistant cells with *RUNX1*, seeded into 12-well plates, and expanded them for 2 days (10^5 cells per plate, $n = 3$). We then re-plated 10^4 of the GFP⁺ puromycin-resistant cells on serum-free E4EC vascular niche layer in haematopoietic media and quantified the number of haematopoietic clusters that emerge after ~20 days of co-culture. We found that GFP⁺ puromycin-resistant cells yielded 156.0 ± 3.6 ($n = 3$) haematopoietic-like colonies per 10^4 re-plated cells, suggesting

that the efficiency of reprogramming was at least 1.5% (156 of 10^4). This calculation assumes that each colony originates from a single reprogrammed cell. The efficiency is probably much higher in cells expressing the appropriate stoichiometric quantities of each factor.

Identification of viral integration on a single-cell and single-colony level. To identify the presence of viral integration on a single-cell level, we sorted human CD45⁺ cells from the marrow of rEC-hMPP engrafted mice into a 96-well plate (1 cell per well) containing a lysis buffer for the Phi29 (multiple displacement amplification; MDA) based whole-genome amplification (WGA). To do single-cell WGA, we used a commercially available kit, REPLI-g (Qiagen, catalogue no. 150343). Each WGA reaction product was followed by a set of PCR reaction with primers specific to the CMV promoter and a transgene (*FOSB*, *GFI1*, *RUNX1*, *SPI1*). All PCR reactions were conducted separately. We used empty wells (no cells sorted) as controls for nonspecific amplification. WGA products of the control wells were used for PCR reactions with primers specific to the CMV promoter and a transgene.

To identify the presence of viral integration on a single-cell level, we captured expanding colonies from the plates for CFC assay. Fourteen days after the start of CFC assay 3 distinct cell aggregations/colonies were detected and analysed. Four PCR reactions were performed for each amplified colony using their genomic DNA as template. Cells from the colonies were re-suspended and washed twice in excessive amounts of PBS (10 ml) and transferred into the lysis buffer for the WGA. All following procedures were the same as those described for the single-cell viral integration identification.

CMV primer, 5'-CGCAATGGGCGGTAGGCGTG-3'; *FOSB* primer, 5'-GC TCTGCTTTTCTCTCTCAACT-3'; *GFI1* primer, 5'-CCAGGGCCCCACAC GGTCGGTAGC-3'; *RUNX1* primer, 5'-TTGCGGTGGGTGTGAAGAC-3'; *SPI1* primer, 5'-CGGATCTTCTTCTGCTGCTGTC-3'.

Clonal reprogramming of HUVECs to rEC-hMPPs. HUVECs were isolated from umbilical cord and grown in endothelial cell growth medium. After 2–3 passages, CD144⁺ CD31⁺ CD62E (E-selectin)⁺ CD45⁺ HUVECs were FACS sorted into 96-well plates at 1, 2, 5 and 10 cells per well densities for clonal expansion. We used CD62E (E-selectin) surface marker to sort mature activated endothelial cells. Passaging of HUVECs results in upregulation of E-selectin in 40–60% of the HUVECs. Expanding clonal populations of selected cells were subsequently transduced with the FGRS transcription factors followed by re-plating onto the E4EC monolayers to reprogram them into rEC-hMPPs. Haematopoietic activity of clonally derived CD45⁺ CD34⁺ rEC-hMPPs was assessed using standard methylcellulose-CFC assay.

RNA-seq processing and analysis. Total RNA was prepared using the Applied Biosystems Artcurus PicoPure RNA isolation kit. The quality of the extracted RNA was checked on an Agilent Technologies 2100 Bioanalyzer. The extracted RNA was used for sequencing using Illumina HiSeq2000. The sequencing output was checked for quality using Illumina pipeline. PE 51x2 and SE 51 reads were mapped to the human genome (hg18) using TopHat (<http://tophat.cbcb.umd.edu/>) default parameters. RefSeq transcript levels, identified as fragments per kilobases of transcripts per million of mapped reads (FPKMs), were then quantified using CuffLinks (<http://cufflinks.cbcb.umd.edu/>) with upper-quartile normalization and sequence-specific bias correction. For heat-map visualization we determined the maximum FPKM of each transcript across the samples shown. FPKMs were then divided by this number to produce scaled expression values. Heat maps of gene expression and gene expression clustering were built using GENE-E matrix visualization and analysis platform (<http://www.broadinstitute.org/cancer/software/GENE-E/>). Clustering of gene expression in the heat maps was conducted using one minus Pearson correlation as dissimilarity measure between transcription profiles. GEO accession number GSE57662.

Comparative genomic hybridization (CGH). Genomic DNA was extracted from HUVECs, FACS sorted CD45⁺ rEC-hMPPs and hCD45⁺ hCD34⁺ rEC-hMPPs sorted from the bone marrow of NSG mice. Before DNA extraction, hCD45⁺ hCD34⁺ rEC-hMPPs sorted from the bone marrow were expanded for 72 h *in vitro*. As a positive control of chromosomal rearrangements we used a CGH array of a leukaemic cell line with a duplication of chromosome 7 and a deletion of chromosome 10. Extracted DNA was digested, labelled by random priming and hybridized to the Agilent 1M CGH arrays. The arrays were scanned in an Agilent DNA microarray scanner and obtained data was visualized using Feature Extraction software (version 10.7; Agilent).

Differentiation and reprogramming of human embryonic stem cells (hES). We used a transgenic hES reporter line that specifically identifies differentiated endothelial cell derivatives via a fluorescent reporter driven by a fragment of the human VE-cadherin promoter^{1,35}. To augment endothelial commitment, hES differentiation was initiated in co-culture with E4EC vascular niche cells, described above. One day before plating hES to begin differentiation, MEF conditioned medium was replaced with hES culture medium without FGF-2 and supplemented with 2 ng ml^{-1} BMP4. The next day, hES cells were plated directly onto E4EC monolayers in hES culture medium (without FGF-2, plus 2 ng ml^{-1} BMP4) and left undisturbed

for 48 h. This point of culture was considered as differentiation day zero. Cells were sequentially stimulated with recombinant cytokines in the following order: day 0 to 7, supplemented with 10 ng ml^{-1} BMP4; day 2 to 14, supplemented with 10 ng ml^{-1} VEGF-A; day 2 to 14, supplemented with 5 ng ml^{-1} FGF-2; day 7 to 14, supplemented with $10 \mu\text{M}$ SB-431542. At day 14 of culture, FACS sorting was used to purify the fraction of hES-derived endothelial cells co-expressing the vascular-specific CD144 (VE-cadherin) reporter and CD31. These cells were transduced with the FGRS cocktail and 2–3 days later plated on a layer of serum-free E4EC monolayers. The extent of reprogramming was assessed by flow cytometry. To accurately detect the expression of CD144 in the hES-ECs being reprogrammed into putative rEC-hMPPs, we used fluorescent monoclonal antibodies to human CD144.

Phagocytosis assay. The rEC-hMPPs generated from 3 to 4 weeks were cultured in the presence of M-CSF (10 ng ml^{-1}), SCF (10 ng ml^{-1}), Flt-3 (10 ng ml^{-1}), TPO (10 ng ml^{-1}) and 10% FBS for an additional 2 weeks with E4EC vascular niche layer. We observed an increase in size and granularity of the cultured cells (data not shown). The culture was washed with PBS twice to remove non-adherent cells. Growth media mixed with green fluorescent beads (GFB) at a low concentration of $1 \mu\text{l ml}^{-1}$ was applied to the attached cells for one hour at 37°C . After the incubation, the cells were washed twice with PBS and live cells were stained with the monocytic CD14 antibody. Cells were fixed and stained with DAPI for nuclear visualization. We visualized GFB inside CD14⁺ cells, but not in CD144 (VE-cadherin)⁺ endothelial cells (Extended Data Fig. 2g).

Purification of human cord blood stem and progenitor cells (HSPCs). Human umbilical cord blood was obtained under the IRB protocol 'stage specific differentiation of hematopoietic stem cells into functional hemangiogenic tissue' (Weill Cornell Medical College IRB 09060010445). Cord blood mononuclear cells were purified by density gradient using Ficoll-Paque (GE) and enriched for CD34⁺ HSPCs using magnetic separation using anti-CD34 microbeads (Miltenyi) or FACS sorting. Further purification was achieved by negative selection of Lin⁺ cells using the human progenitor cell enrichment kit (StemCell Technologies) or FACS sorting. RNA from FACS sorted Lin[−]CD34⁺CD45⁺ cells was isolated by using Arcturus PicoPure RNA isolation kit (Applied Biosystems; this kit was used for all RNA extraction procedures).

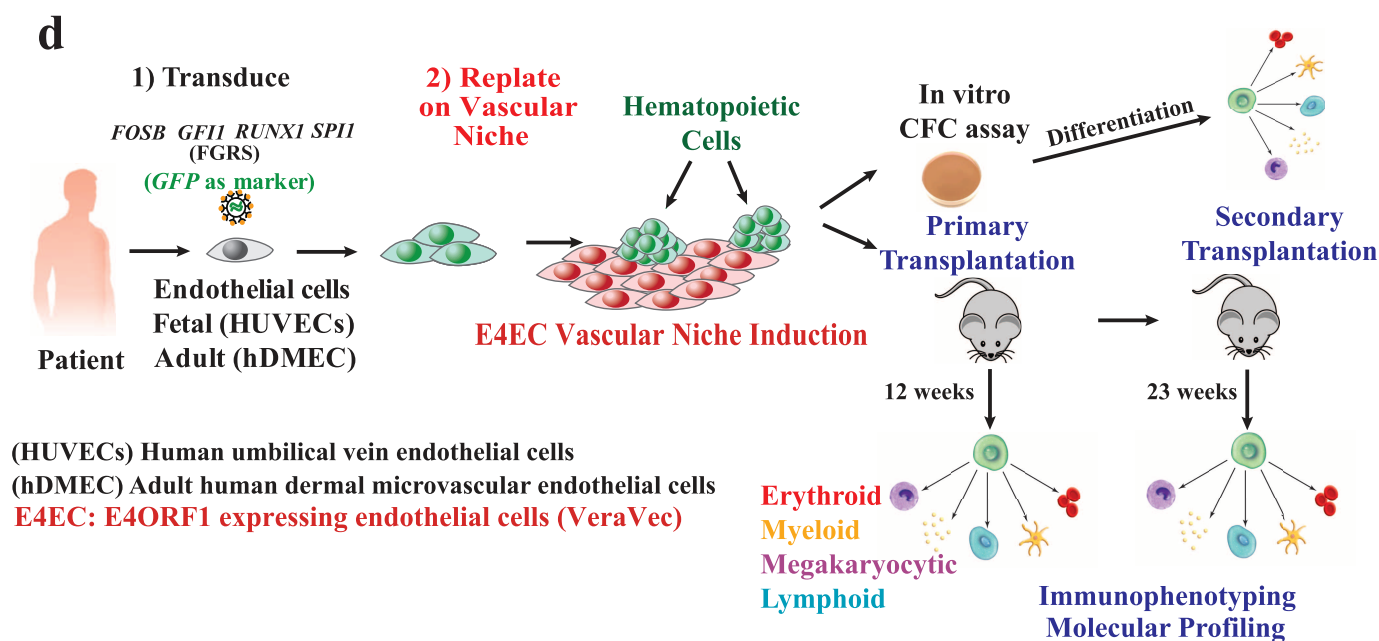
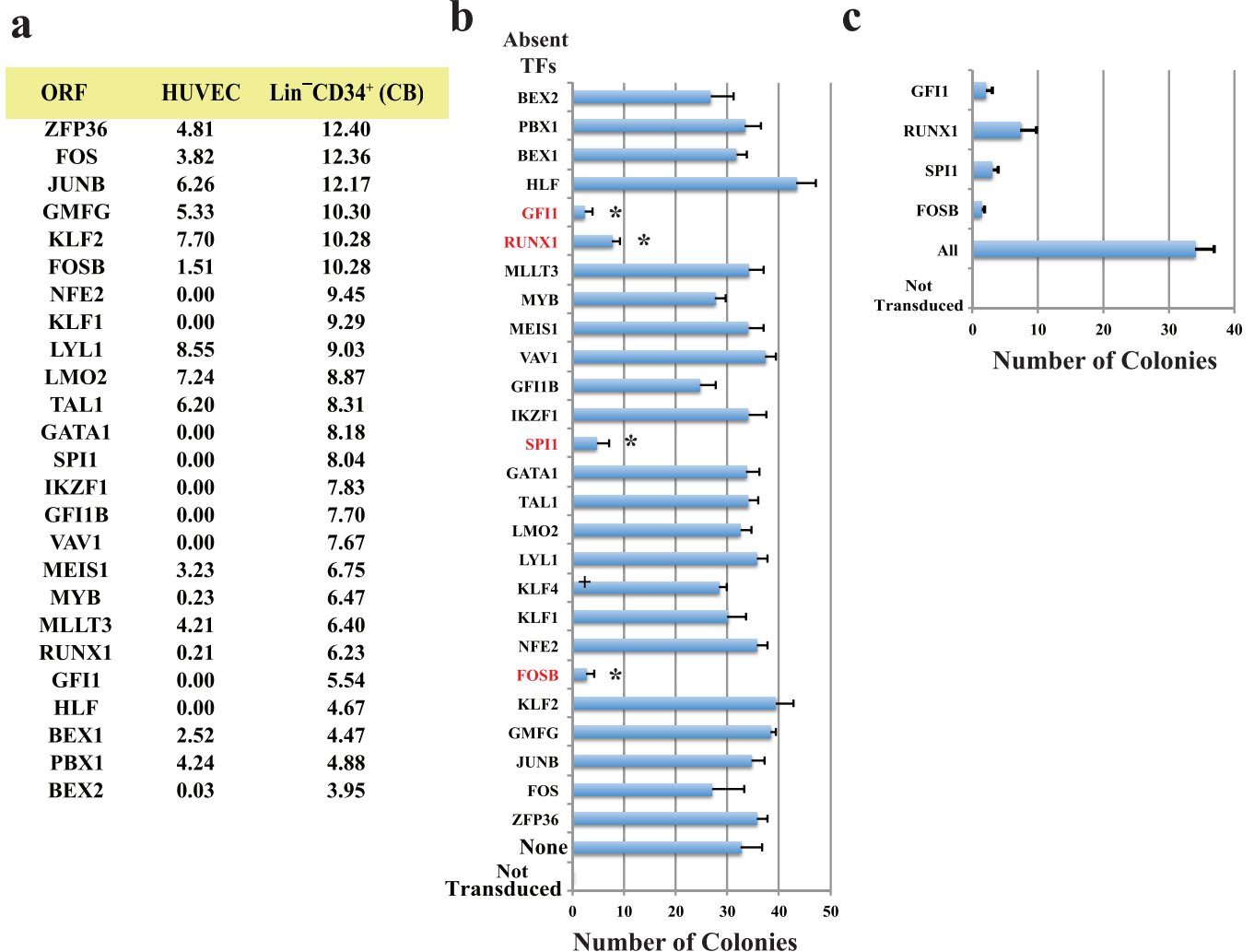
Lentiviral vectors. Candidate transcription factors used for screening were subcloned into pLVX-IRES-ZsGreen1 lentivector (Clontech), pLOC lentivector (OpenBiosystems), or LV105 and LV151 lentivectors (Genecopoeia). pLOC lentiviral vector contained CMV-MCS-IRES-GFP (MCS, multicloning site where a cDNA of interest such as *FOSB* or *GFI1* was subcloned). Human FGRS were subcloned as follows: *FOSB* and *GFI1* were each subcloned into pLOC lentivector containing IRES-GFP cassette, SPII was subcloned into LV105 lentivector containing puromycin selection marker, and *RUNX1* was subcloned into LV151 lentivector containing neomycin selection marker. Tet-On 3G inducible lentivectors (Clontech) were used for inducible expression of either mouse FGRS (mFGRS) or human FGRS (hFGRS) factors. Expression of all FGRS transgenes was driven by the CMV promoter. Lentiviral particles were packaged as previously described⁷. In short, human embryonic kidney 293FT (HEK293FT) cells were co-transfected with a lentiviral vector and two helper plasmids, psPAX2 and psMD2.G (Trono Lab through Addgene), in an equal molar

ratio. Supernatant was collected 48–52 h after transfection, filtered and concentrated using Lenti-X concentrator (Clontech). Viral titres were determined in limiting dilution experiments using HUVECs as target cells. We used either the number of GFP⁺ cells or the number of formed colonies in the presence of selection antibiotics (puromycin) as a read-out for the number of infectious viral particles per volume. We used an average multiplicity of infection (MOI) of 5 to 10 for infection of endothelial cells.

Flow cytometry. Flow cytometry analysis was performed on a Becton Dickinson LSRII SORP, and FACS was performed on an Aria II SORP. Antibodies used were raised against human CD45, CD34, CD14, CD31, CD43, CD90, CD41a, CD33, CD19, CD3, CD4, CD8, CD235, CD45RA, CD83, CD11b, CD38, Lin cocktail, CD117, CD133, CD144 (BD Pharmingen, eBioscience) or mouse CD45 (eBioscience.) Voltage adjustments and compensation was performed with CompBeads (BD Pharmingen), and gating was performed on fluorophore minus one (FMO) controls and unstained controls.

The list of antibodies used in our experiments is given below. Anti-human antibodies obtained from eBioscience: CD45 catalogue no. 47-0459-42; clone HI30, CD34 catalogue no. 25-0349-42; clone 4H11, CD33 catalogue no. 48-0337-42; clone p67.6, CD19 catalogue no. 12-0199-41; clone HIB19, CD3 catalogue no. 93-0037-42; clone OKT3, CD4 catalogue no. 17-0048-41; clone OKT4, CD8 catalogue no. 8048-0087-025; clone SK1, CD43 catalogue no. 17-0439-73; clone eBio84-3C1, CD83 catalogue no. 25-0839-41; clone HB15e, CD11b catalogue no. 12-0118-41; clone ICRF44, LIN catalogue no. 22-7778-72, CD31 catalogue no. 11-0319-42; clone WM59, CD31 catalogue no. 48-0319-42; clone WM59. Anti-human antibodies obtained from BD Pharmingen: CD90 catalogue no. 561971; clone 5E10, CD3 catalogue no. 557851; clone SK7, CD14 catalogue no. 557742, CD14 catalogue no. 555399, CD235A catalogue no. 340947, CD45RA catalogue no. 347723, CD41a catalogue no. 555466, CD38 catalogue no. 646851, CD117 catalogue no. 333944, CD33 catalogue no. 333946, CD144 catalogue no. 560410; clone 55-7H1, FLK1 (VEGF-R2) catalogue no. 560871; clone 89106. Anti-human antibodies obtained from BioLegend: Lin catalogue no. 348805. Anti-mouse antibodies obtained from eBioscience: CD45 catalogue no. 25-0451-82; clone 30-f11.

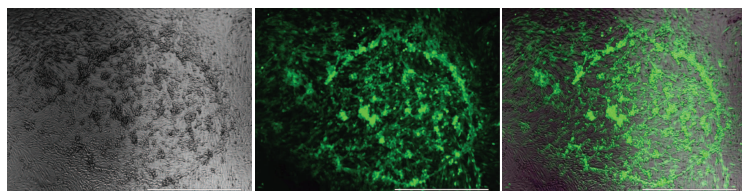
Statistics and animals. All statistics are presented as average \pm standard deviation. To identify statistical significance all groups of data were compared using a paired student *t*-test. Experiments were repeated for at least three times. Number of repeats is demonstrated in all figure legends. Animal experiments contain at least three animals per group. The number of animals is described in all figure legends and the text of the paper. We included all tested animals for quantification. Representative images and flow cytometry plots are shown in the figures. Age- and sex-matched animals were allocated in all corresponding experimental groups. All NSG (NOD.Cg-Prkdc^{scid}Il2rg^{tm1Wjl}/SzJ, Jackson laboratory) animals for transplantation experiments were female. All ages are specified in the text. Animals were chosen according to their age and their sex (females only). A description of every experiment states the age of the animals used in the experiment. Transplanted animals were not individually labelled. Hence, subgroups of transplanted animals for organ engraftment were chosen blindly, without previous knowledge of the level of engraftment. Animal experiments were performed under the guidelines set by the Institutional Animal Care and Use Committee (IACUC).



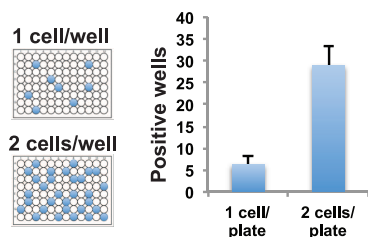
Extended Data Figure 1 | Screening strategy for identification of a minimal set of transcription factors for reprogramming endothelial cells into haematopoietic cells. **a**, Candidate genes tested for reprogramming of HUVECs into haematopoietic colonies. To identify transcription factors that drive EHT transition, we performed RNA-seq of HUVECs and Lin[−]CD34⁺ umbilical cord HSPCs to select transcription factors differentially expressed by HSPCs, but not by HUVECs. We then screened various combinations of differentially expressed transcription factors to identify a minimal set capable of reprogramming endothelial cells to haematopoietic cells. Levels of expression (\log_2 [RNA-seq value]) in HUVECs and freshly purified Lin[−]CD34⁺ cord blood cells are shown. **b**, One-by-one elimination of transcription factor experiment revealed a minimal set of transcription factors—*FOSB*, *GFI1*, *RUNX1* and *SPI1* (FGRS)—capable of generating haematopoietic colonies in the HUVEC culture. A pooled set of 26 transcription factors minus one transcription factor was evaluated for the ability to evoke formation of haematopoietic clusters (day 21–25; $n = 3$). Asterisks show statistically

significant ($P < 0.05$) reduction of the number of haematopoietic clusters in the transduced HUVECs compared to the full set of transcription factors. Control represents non-transduced HUVECs. Transduced cells were cultured on a layer of serum-free E4EC monolayers. **c**, One-by-one elimination of the FGRS factors shows that all four FGRS transcription factors are necessary and sufficient for generation of haematopoietic colonies (day 21–25; $n = 3$). ‘All’ in **b** and **c** indicates that all transcription factors are present. ‘Not transduced’ in **b** and **c** indicates that all transcription factors are absent. **b**, **c**, Error bars are average \pm s.d. **d**, Schema for reprogramming of endothelial cells into human multipotent progenitor cells (rEC-hMPPs). Clonal or bulk populations of HUVECs or hDMECs were transduced with the FGRS and after 3 days were replated on subconfluent monolayers of E4EC endothelial cells (VeraVecs). The emerging colonies of haematopoietic cells were subjected to (1) multivariate immunophenotypic analyses; (2) clonal and oligo-clonal CFC assay; and (3) molecular profiling (RNA-seq). Tissues of the engrafted mice were processed for histological examination to rule out any malignant transformation.

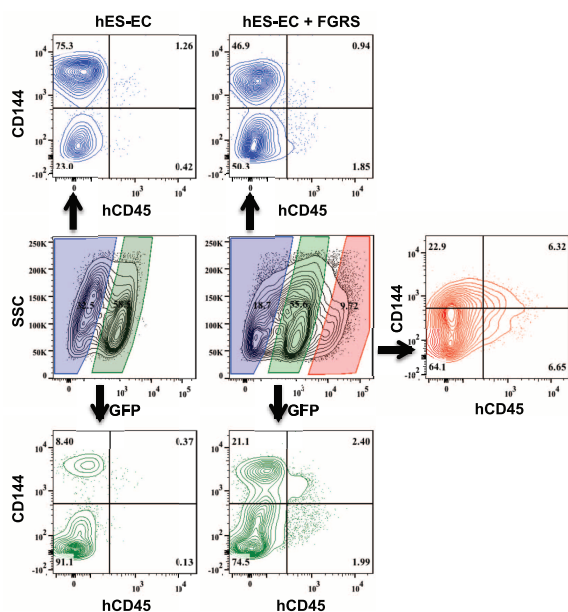
a



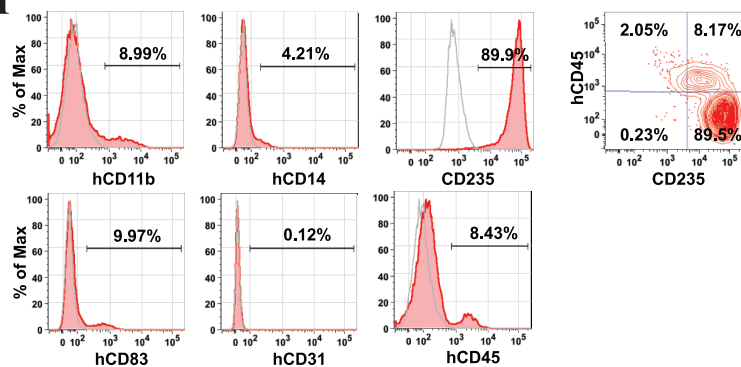
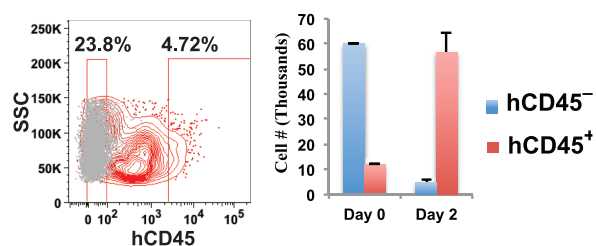
c



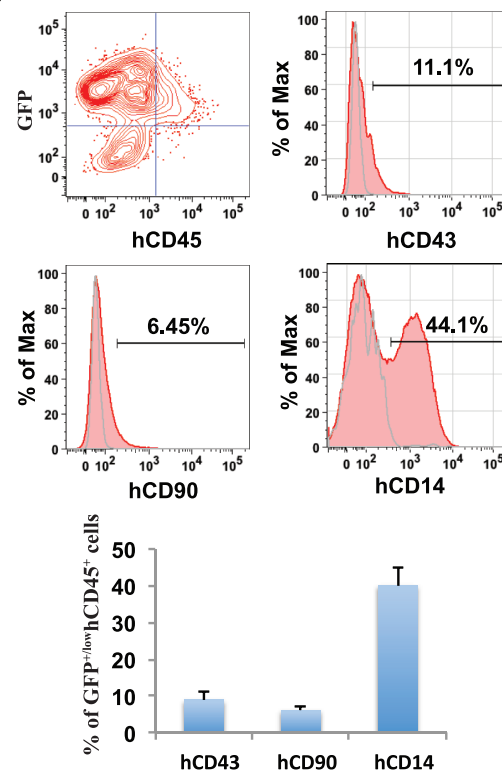
d



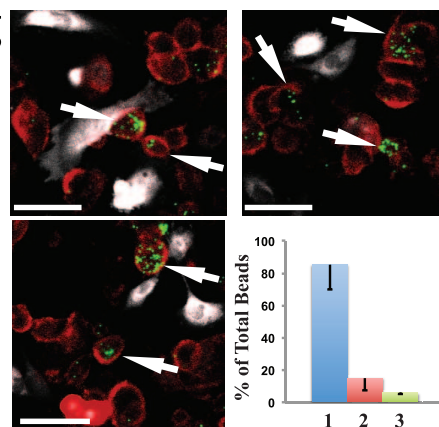
f

**b**

e



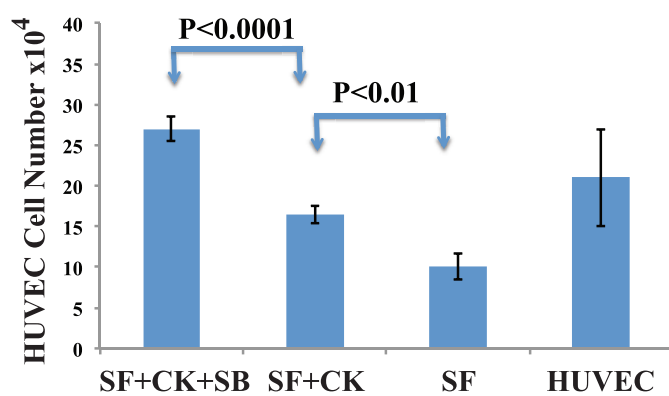
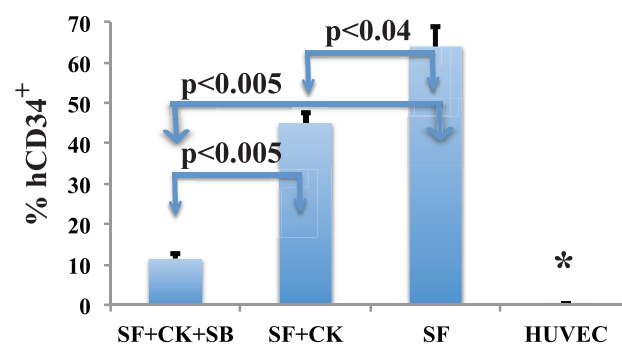
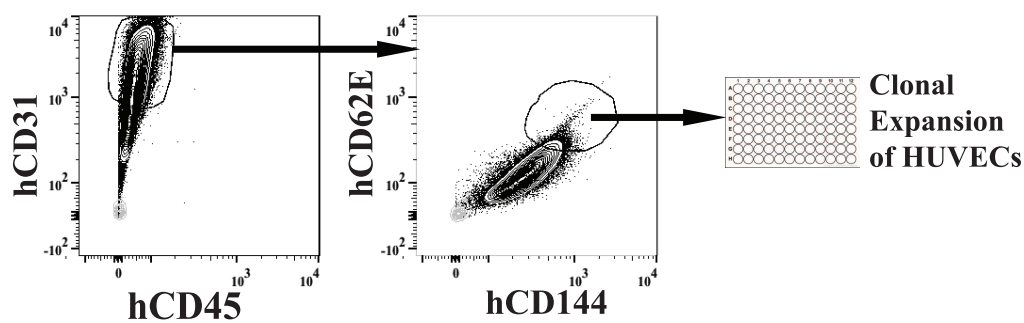
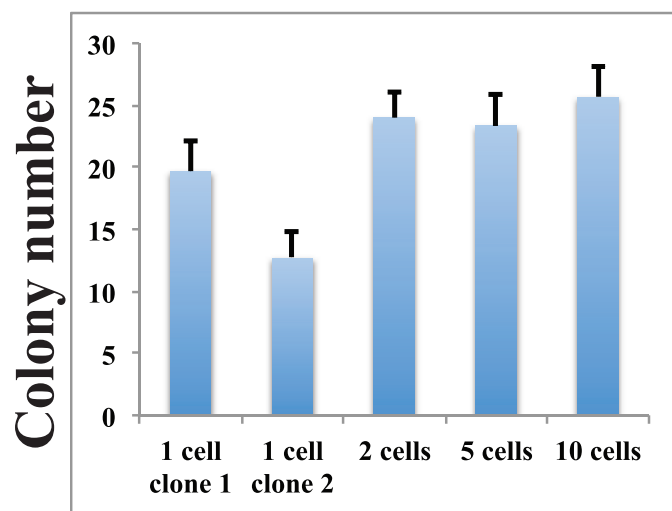
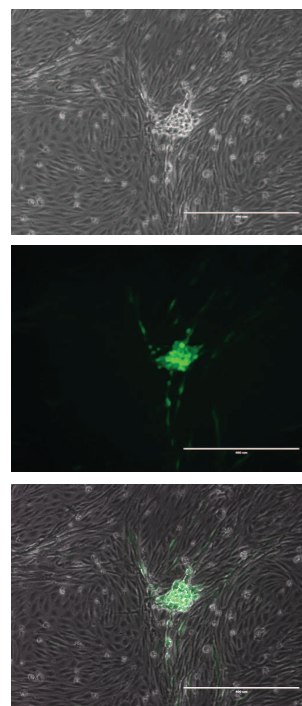
g



Extended Data Figure 2 | FGRS transduction and vascular induction reprogram HUVECs, but not hES-ECs, to proliferating functional rEC-hMPPs.

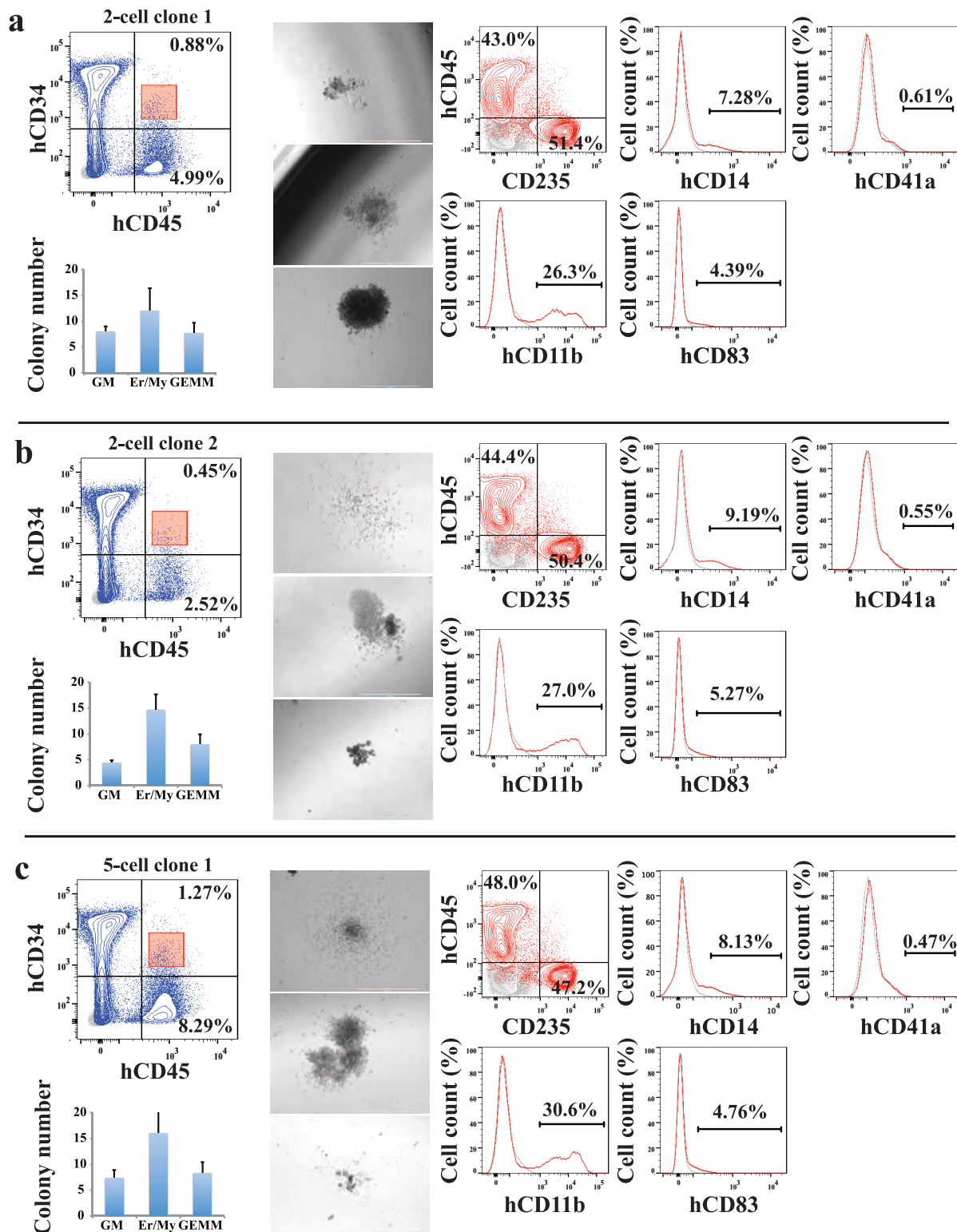
a. Multi-colony niche-like structure that physically separates developing haematopoietic colonies from surrounding E4EC vascular niche. The emerging multi-colony sinusoidal-like structures create a unique cellular interface between E4EC monolayers and transduced endothelial cells giving rise to haematopoietic clusters ($n = 4$, scale bar is 1,000 μm). **b.** Expansion potential of reprogrammed hCD45⁺ haematopoietic cells. hCD45⁺ (12×10^3) and hCD45⁻ (60×10^3) cells were sorted into separate wells and expanded for 2 days. We observed fivefold expansion of hCD45⁺ cells ($56.6 \times 10^3 \pm 7.9 \times 10^3$; $n = 3$) and marked reduction of hCD45⁻ cell number ($4.6 \times 10^3 \pm 1.0 \times 10^3$; $n = 3$). **c.** Clonal expansion of hCD45⁺ cells. hCD45⁺ cells were FACS sorted into 96-well plates at the density of 1 and 2 cells per well. After 7 days of culture, we observed hCD45⁺ cell expansion in 6.3 ± 2.1 wells (93.1 ± 14.5 cells per well) of 1-cell sort and 29.0 ± 4.3 wells (112.1 ± 21.2 cells per well) of the 2-cell sort ($n = 3$). The difference between cell number in 1- and 2-cell sort was statistically not significant ($P = 0.78$), suggesting that the difference in the number of wells with detected cell expansion was due to survival of sorted cells rather than a reflection of the number of cells sorted into a well. **d.** Reprogramming of hES-derived endothelial cells (hES-ECs) into haematopoietic cells. Representative experiment demonstrating that transduction of hES-ECs with FGRS (F and G lentivector constructs containing IRES-GFP cassette) and E4EC vascular induction generated significantly

higher numbers of GFP⁺hCD45⁺hCD144⁻ cells (four panels on the right) compared to control non-transduced hES-ECs (three panels on the left). To accurately detect the expression of CD144 in the hES-ECs being reprogrammed into putative rEC-hMPPs, we used fluorescent monoclonal antibodies to human CD144. **e.** Lineage-specific surface marker analysis of the hGFP⁺CD45⁺ population of rEC-hMPPs. hGFP⁺CD45⁺ population showed that some of these cells expressed lineage-specific surface markers, such as hCD43⁺ ($8.96 \pm 2.3\%$; $n = 3$), hCD90⁺ (Thy-1⁺) ($6.15 \pm 1.13\%$; $n = 3$) and hCD14⁺ ($40.0 \pm 4.95\%$; $n = 3$) (representative flow cytometry measurements; top four panels, statistics for all experiments is in the bottom bar graph, $n = 3$). **f.** Immunophenotypic analysis of CFC colonies grown in the CFC assay performed in Fig. 2c, d. **g.** Macrophages differentiated from rEC-hMPPs are functionally capable of phagocytosis. The images (upper row and lower left) show groups of firmly plastic-adherent hCD14⁺ cells (red staining) with clearly visible phagocytosed green fluorescent beads (GFB; green). Endothelial CD144⁺ (VE-cadherin) cells (white staining) were not co-localized with beads. Most ($85.1 \pm 15.1\%$) GFBs were localized inside hCD14⁺ cells (bottom-right graph, 1). A smaller population of GFBs was distributed outside hCD14⁺ and CD144 (VE-cadherin)⁺ cells ($14.8 \pm 7.43\%$; bottom-right graph, 2). The percentage of GFBs co-localized with endothelial cells was negligible ($4.8 \pm 0.83\%$; bottom-right graph, 3), $n = 9$. Scale bars are 25 μm . Error bars are average \pm s.d.

a**b****c****d****e**

Extended Data Figure 3 | Naive HUVECs are devoid of haemogenic potential capable of spontaneous differentiation into MPPs. We performed two sets of experiments to exclude the possibility that rEC-hMPPs were derived from spontaneously differentiating HUVECs with haemogenic or haemangioblastic potential^{25,34}. **a, b**, In optimal pro-haematopoietic cultures, naive non-transduced endothelial cells fail to spontaneously differentiate into rEC-hMPPs. **a**, We grew non-FGRS-transduced HUVECs in the serum-free media used for reprogramming. Neither serum withdrawal nor addition of haematopoietic cytokines induced formation of hCD45⁺hCD34⁺ cells and HUVECs sustained their vascular identity. Indeed, serum withdrawal increases the number of CD34⁺ HUVECs. CK, cytokine cocktail (see Methods); SB, TGF- β inhibitor SB431542; SF, serum free. **b**, Serum withdrawal suppresses HUVEC proliferation. Inhibition of TGF- β signalling (SB) combined with cytokine cocktail (see Methods) restores proliferative potential of HUVECs in serum free media. The difference between proliferation of HUVECs in serum free media and all other conditions is statistically significant (asterisk; $P < 0.005$). Statistical significance between pairs of different conditions is shown with blue arrows and P values, where $P < 0.005$ is statistically significant. Therefore, human rEC-hMPPs originate from reprogrammed endothelial cells,

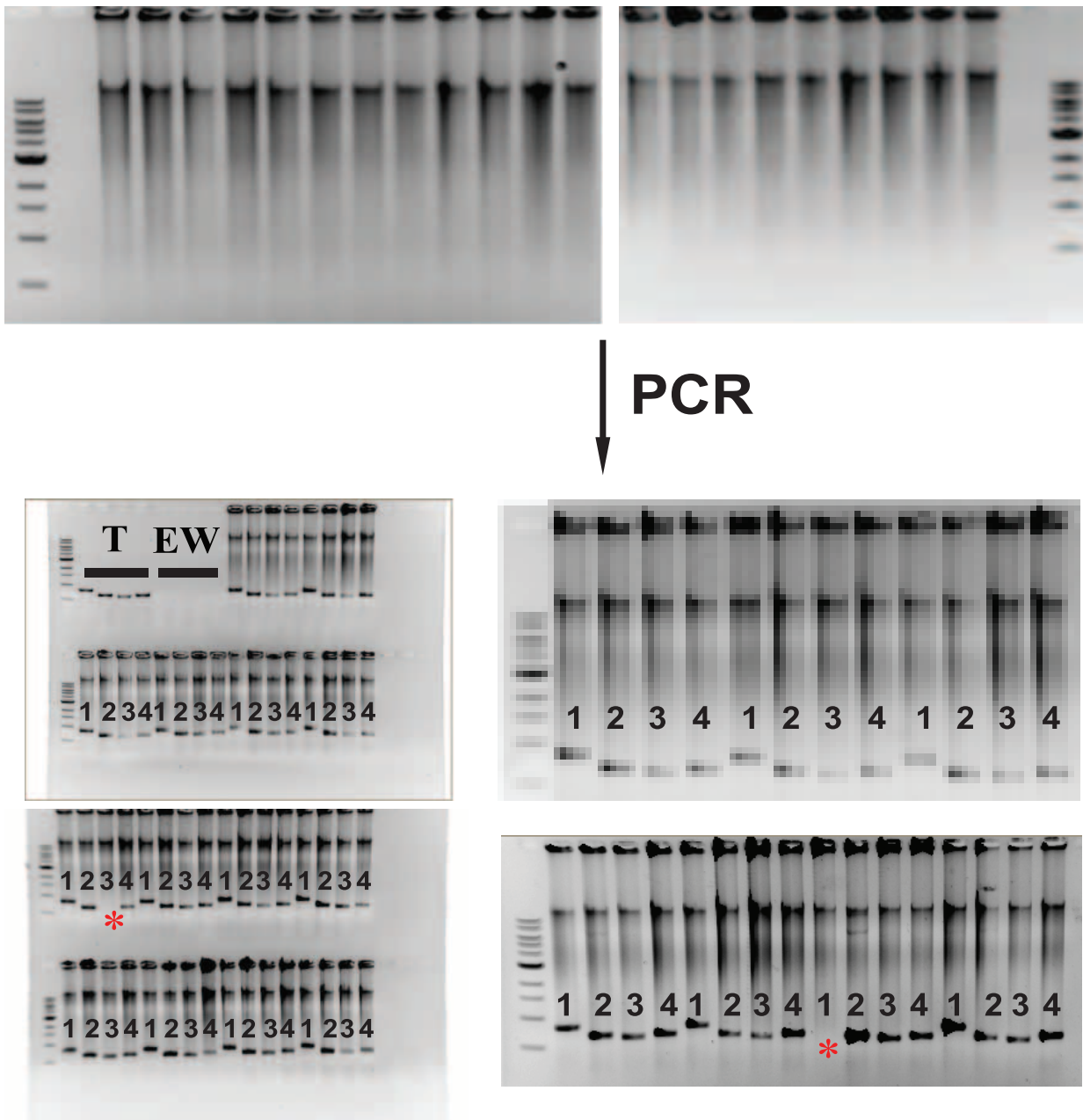
but not cytokine-mediated outgrowth of contaminating pre-existing haemogenic endothelial cells. **c–e**, Clonal reprogramming of non-haemogenic HUVECs into rEC-hMPPs using FGRS transduction and vascular induction. We performed endothelial cell clonal reprogramming experiments to exclude the possibility that rEC-hMPPs were derived from spontaneously differentiating HUVECs with pre-existing haemogenic or haemangioblastic potential^{25,34}. **c**, Because E-selectin is only expressed in activated endothelial cells, we generated clonal cultures of CD45[−]CD144⁺CD31⁺CD62E[−] (E-selectin)⁺ endothelial cells^{32,33}. To this end, CD144⁺CD31⁺CD62E⁺CD45[−] HUVECs were sorted into 96-well plates at 1, 2, 5 and 10 cells per well densities for clonal expansion. Proliferating clones were transduced with FGRS and induced with serum-free E4EC monolayers. **d**, These clonal cultures yielded rEC-hMPPs comparable to bulk HUVEC cultures. The numbers of haematopoietic-like colonies emerging from 1-cell, 2-cell, 5-cell and 10-cell clones are not statistically different ($P > 0.05$). **e**, An example of a haematopoietic-like colony derived from a 1-cell clone number 2. It is unlikely that rEC-hMPPs are derived through spontaneous differentiation of pre-existing endothelial cells with haemogenic or haemangioblastic potential. Error bars are average \pm s.d. Scale bars, 400 μ m.



Extended Data Figure 4 | Clonal reprogramming of non-haemogenic HUVECs into immunophenotypic and functional rEC-hMPPs using FGFR3 transduction and vascular induction. **a–c**, CFC assay of reprogrammed hCD45⁺hCD34⁺ rEC-hMPPs generated from clonally selected CD45⁺CD144⁺CD31⁺CD62E⁺ mature HUVECs, as shown in Extended Data Fig. 3d, e. CD45⁺CD144⁺CD31⁺CD62E⁺ endothelial cells were sorted as 1 cell per well, 2 cells per well, and 5–10 cells per well. Expanding clones of the endothelial cells were transduced with FGFR3 and then induced with vascular niche. After 3 to 4 weeks, emerging hCD45⁺hCD34⁺ rEC-hMPPs were sorted out (red gate in FACS plots; upper left) and plated for CFC assay. Typical haematopoietic colonies arose in the assay (middle column,

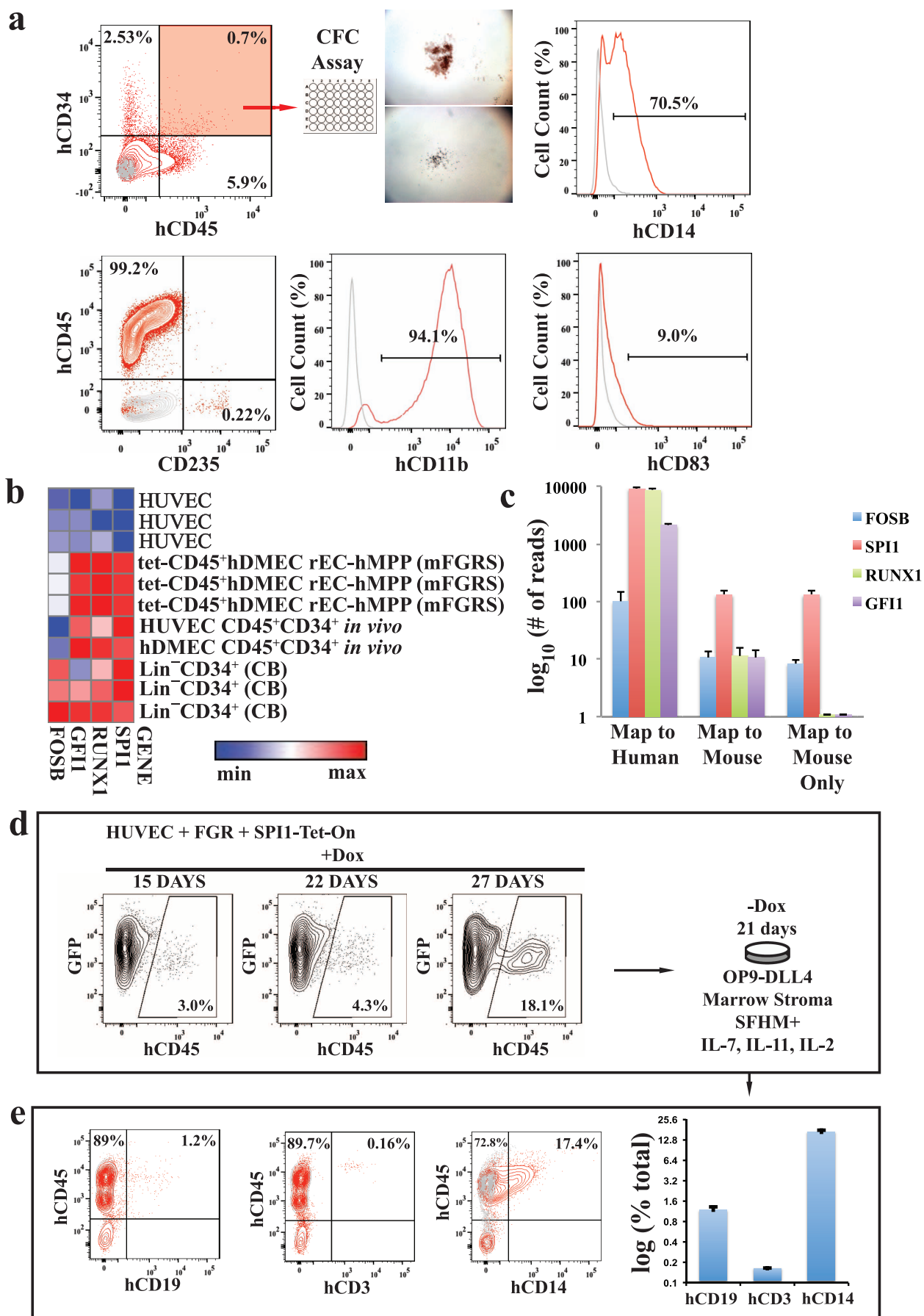
microphotographs; original magnification, $\times 4$). FACS plots on the right show immunophenotypic analysis of the cells that arose in the CFC assay, demonstrating differentiation into human CD45⁺CD235⁺ erythroid, CD11b⁺ macrophage CD14⁺ monocytic, CD41a⁺ megakaryocytic and CD83⁺ dendritic progenies. The graph in the left lower corner shows quantification of the CFC assay ($n = 3$). Identical panels are shown for two 2-cell clones and one 5-cell clone. A total of three independent clones is shown. Thus, given the high efficiency of clonal reprogramming of mature authentic endothelial cells into rEC-hMPPs, it is unlikely that rEC-hMPPs are spontaneously derived from a very rare population of a pre-existing haemogenic or haemangioblastic HUVECs. Scale bars, 400 μ m.

WGA (21 cells)



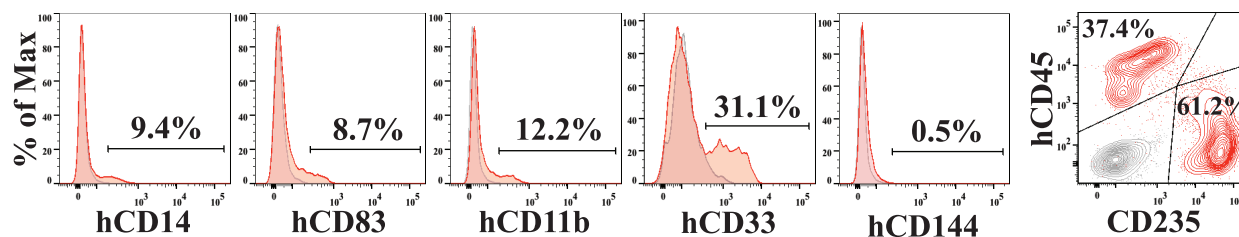
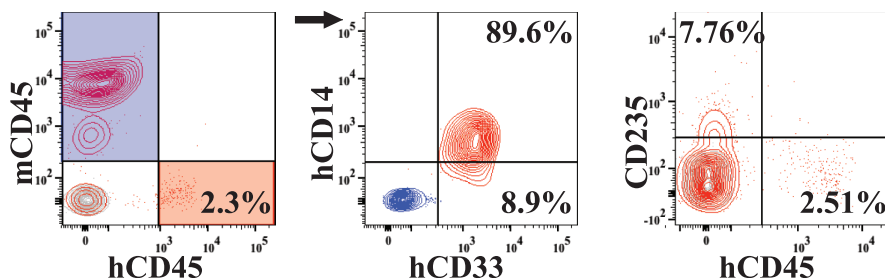
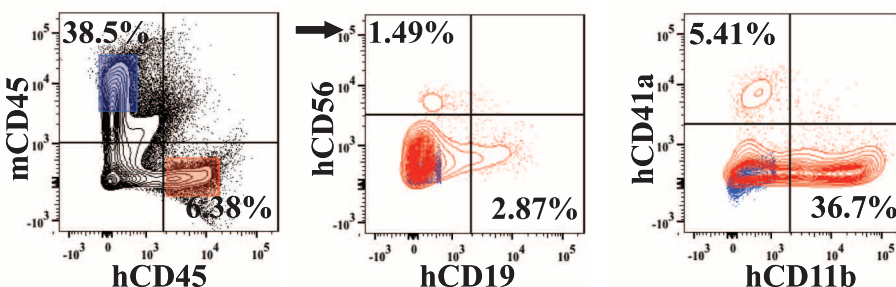
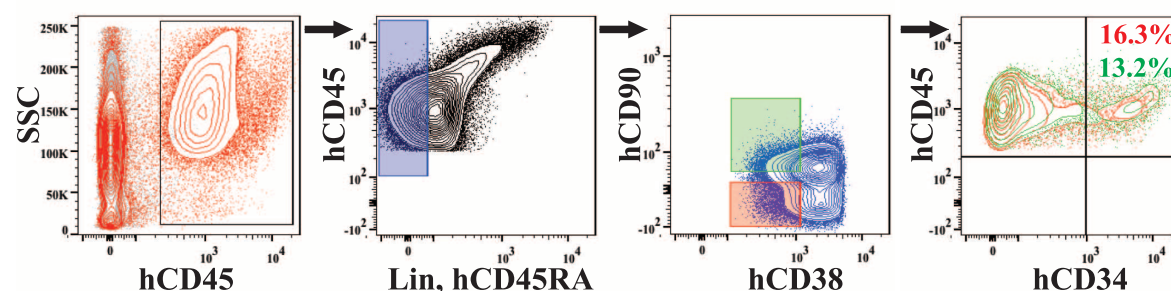
Extended Data Figure 5 | Single-cell analysis of lentiviral integration into engrafted rEC-hMPPs. Engrafted hCD45⁺ rEC-hMPPs purified from the bone marrow of primary NSG recipient mouse (Fig. 3e, f) were sorted into a 96-well plate (1 cell per well), lysed in corresponding well for whole genome amplification (WGA) using Phi29 enzyme (see Methods). Amplified DNA is shown for all 21 cells in the top two gels. Amplified DNA was used as a template

for PCR reactions with a forward primer specific for the CMV promoter and reverse primer specific for the coding sequence of a reprogramming factor. *t*-test PCR with a lentiviral vector. EW indicates empty well (no template DNA). Red asterisks show failed PCR amplification of viral integration. PCR products are visible as low molecular mass bands labelled as 1, *FOSB*; 2, *GFI1*; 3, *RUNX1*; 4, *SPI1*.



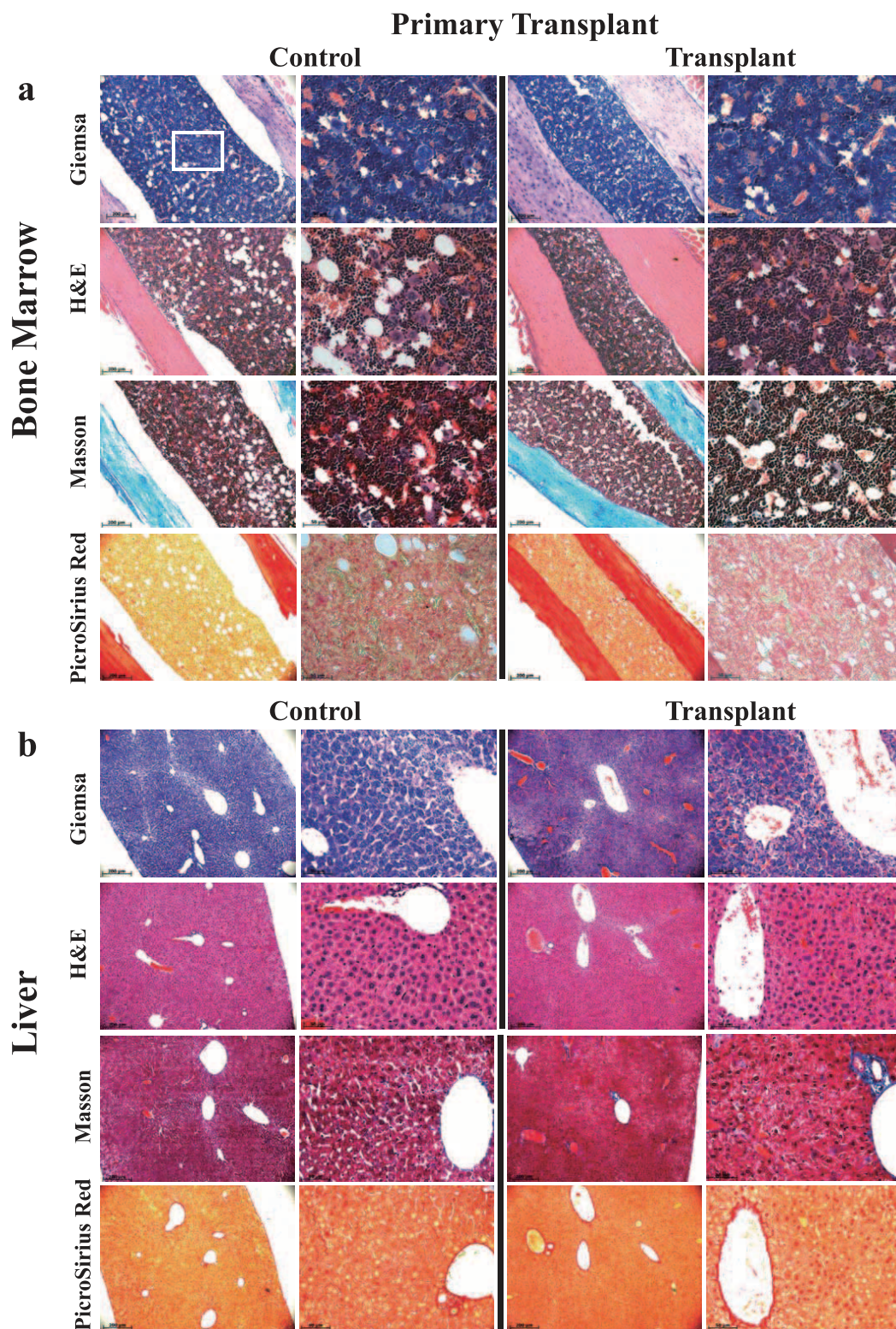
Extended Data Figure 6 | Conditional expression of FGRS is sufficient for optimal generation of rEC-hMPPs with multilineage potential, including T-cell lymphoid cells. **a–c**, Conditional expression of mouse inducible FGRS (mFGRS) factors activates endogenous human FGRS in HUVECs sustaining functional haematopoietic cell fate of rEC-hMPPs. **a**, To test whether FGRS-induced reprogramming triggered expression of endogenous FGRS genes²⁴, HUVECs were transduced with lentiviral vectors expressing mFGRS-Tet-On and a trans-activator, and grown on E4EC vascular niche for 18–22 days ($n = 4$) in the presence of doxycycline. Doxycycline was removed from the culture medium after 18–22 days to shut off the expression of mouse FGRS and cells were cultured for an additional 7–10 days. Human CD45⁺CD34⁺ cells were FACS isolated for CFC assay and whole-transcriptome deep sequencing (RNA-seq). CFC assay revealed emergence of haematopoietic colonies with cells expressing human CD235, CD11b, CD83 and CD14. **b**, Comparison of transcriptional gene profiles of human FGRS in: primary HUVECs; rEC-hMPPs reprogrammed from hDMECs by transduction with inducible Tet-On mouse FGRS (mFGRS) and vascular induction for 3–4 weeks (tet-CD45⁺hDMEC-rEC-hMPPs); engrafted human CD45⁺CD34⁺ cells purified from the bone marrow of NSG mice 22 weeks post-primary transplantation with HUVEC-reprogrammed rEC-hMPPs (HUVEC CD45⁺CD34⁺ *in vivo*); engrafted human CD45⁺CD34⁺ cells purified from the bone marrow of NSG mice 15 weeks post-secondary engraftment with hDMEC-reprogrammed rEC-hMPPs (hDMEC CD45⁺CD34⁺ *in vivo*); and naive purified Lin[−]CD34⁺ cells from cord blood (CB). **c**, Analysis of whole-transcriptome RNA-seq of rEC-hMPPs derived using inducible mouse FGRS ($n = 3$). All RNA-seq reads were aligned against human and mouse

FGRS sequences. ‘Map to human’ indicates RNA-seq reads that align to human FGRS sequences; ‘Map to mouse’ indicates RNA-seq reads that align to mouse FGRS sequences; and ‘Map to mouse only’ indicates RNA-seq reads that align to mouse FGRS sequences without a possibility to align to human sequences. **d, e**, Optimizing differentiation of rEC-hMPPs into lymphoid progeny. **d**, The number of T-lymphoid progeny of engrafted rEC-hMPPs was negligibly small, raising the possibility that constitutive SPI1 expression prevents rEC-hMPPs from differentiating into T cells^{39,40}. To test this, HUVECs were transduced with lentiviral vectors expressing GFP and that constitutively express FGR transcription factors with a Tet-inducible *SPI1* (FGR+*SPI1*-Tet-On construct) for 3 days followed by re-plating for E4EC induction. After 27 days of FGR and doxycycline-induced *SPI1* expression on E4ECs, GFP⁺hCD45⁺ haematopoietic-like colonies emerged. Then, doxycycline was withdrawn and the reprogrammed cells were cultured serum-free haematopoietic media (SFHM) with Delta-like-4 expressing OP9-stroma (OP9-DLL4) supplemented with IL-7, IL-11 and IL-2. There is an increase of the number of GFP⁺hCD45⁺ cells emerging during reprogramming of HUVECs by FGR+*SPI1*-Tet-On construct and E4EC induction. **e**, rEC-hMPPs differentiate into CD3⁺, CD19⁺ and CD14⁺ haematopoietic cells in the absence of exogenous expression of *SPI1*. After 3 weeks, the numbers of myeloid and lymphoid cells were quantified by flow cytometry. We were able to reliably detect a small fraction of CD3⁺ cells ($0.16 \pm 0.01\%$; $n = 3$), a larger number of CD19⁺ ($1.17 \pm 0.13\%$; $n = 3$) and CD14⁺ ($16.46 \pm 1.02\%$; $n = 3$) cells. Thus, generation of lymphoid cells from rEC-hMPPs could be optimized by transient expression of transcription factors. Error bars are average \pm s.d.

a**b PB****c Spleen****d BM**

Extended Data Figure 7 | Adult human hMEC-derived rEC-hMPPs are capable of *in vivo* multilineage engraftment. **a**, Immunophenotypic analysis of cells grown in the CFC assay (from Fig. 4b). These panels show quantification of surface marker expression in the cells isolated from colonies in the CFC assay ($n = 3$). hMECs differentiated into hCD45⁺CD235⁺ erythroid, CD11b⁺CD14⁺ monocyte/macrophage and CD83⁺ dendritic cell progenies. Minimal CD144 (VE-cadherin) was detected. **b**, Analysis of peripheral blood (PB) of mice at 4, 6 and 12 weeks after primary transplantation (Fig. 5a) revealed circulating hCD45⁺ and their hCD33⁺, hCD14⁺ myeloid and hCD45⁺CD235⁺ erythroid progenies ($n = 6$). Mouse CD45 (mCD45⁺) cells were excluded from analyses. Mouse cells, blue; human cells, red.

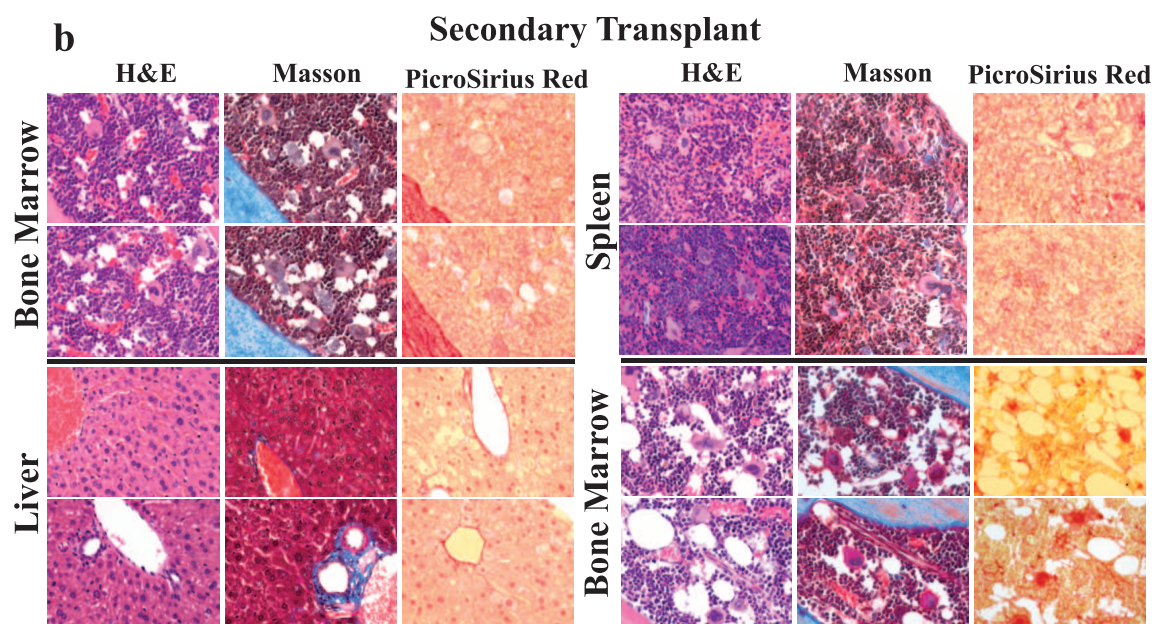
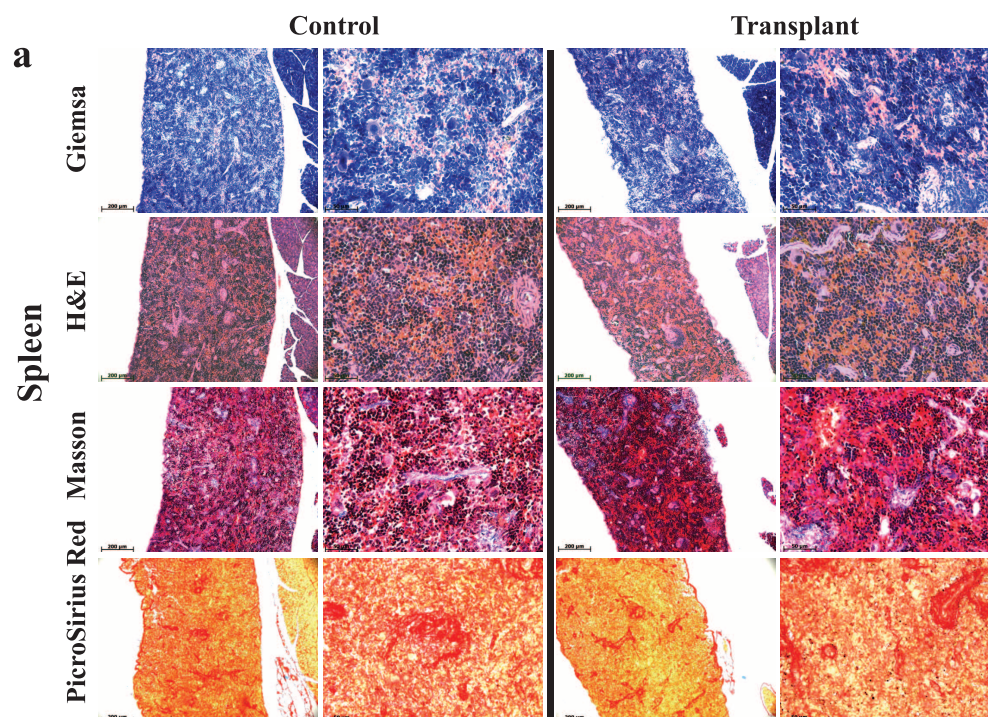
c, Analysis of spleen of mice at 14 weeks after primary transplantation (Fig. 5a) revealed the presence of hCD45⁺ (red gate) and their lymphoid (hCD19⁺, hCD56⁺) and myeloid (hCD11b⁺, hCD41a⁺) progenies ($n = 3$). Mouse CD45 (mCD45⁺ cells, blue populations) is shown. **d**, Analysis of mouse bone marrow at 14 weeks after primary transplantation (Fig. 5a). Lin⁺CD45RA⁺ cells (blue gate) were analysed for CD38 and CD90 expression (green and red gates) and subsequently examined for human CD45 and CD34 expression. This analysis revealed the presence of hCD34⁺ cells with small populations of both Lin⁺CD45RA⁺CD38⁺CD90⁺CD34⁺ and Lin⁺CD45RA⁺CD38⁺CD90⁺CD34⁺ cells satisfying phenotypic definition of human HSCs and MPP, respectively ($n = 3$).



Extended Data Figure 8 | Analysis of bone marrow and liver of primary transplanted mice for signs of malignant transformation. Analysis of bone marrow (a) and liver (b) of mice 10 months after primary transplantation (from Fig. 3b) of HUVEC-derived rEC-hMPPs for signs of malignant transformation. The level of fibrosis was determined using Masson and PicroSirius staining. The architectonic geometry of the bone marrow was determined by sequential multi-cross-sectional Wright–Giemsa and haematoxylin and eosin (H&E) staining and compared to age-controlled, non-transplanted NSG mice. We did not observe any evidence of fibrosis or alteration of the geometry of the bone marrow or liver of the transplanted mice. Furthermore, no recipient mouse manifested any anatomical or symptomatic

evidence of leukaemias, lymphomas or myeloproliferative neoplasm (MPN) (that is, lymphadenopathy, organomegaly, illness or haemorrhage). Circulating hCD45⁺ cells in peripheral blood displayed no evidence of lympho/myeloproliferation or dysplasia. Furthermore, microscopic architecture of bone marrow and liver was normal and without fibrotic remodelling or aberrant deposition of collagen or desmin. All images were acquired using a colour CCD camera. The scale bar is 200 μ m for low-resolution images in the left columns and 50 μ m for high-resolution images in the right columns. Upper-left image (Giemsa, control) shows a white square in the centre that corresponds to the portion of the image shown at high resolution on the right (the same Giemsa control sample). This rule applies to all images shown.

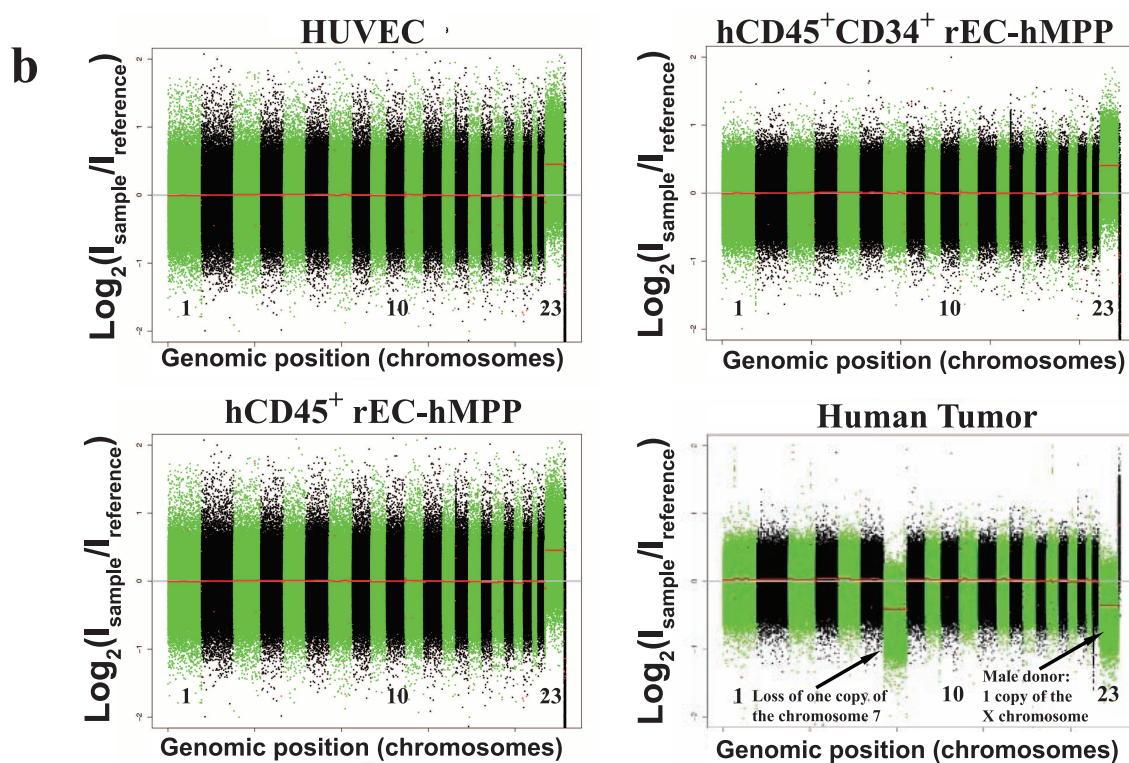
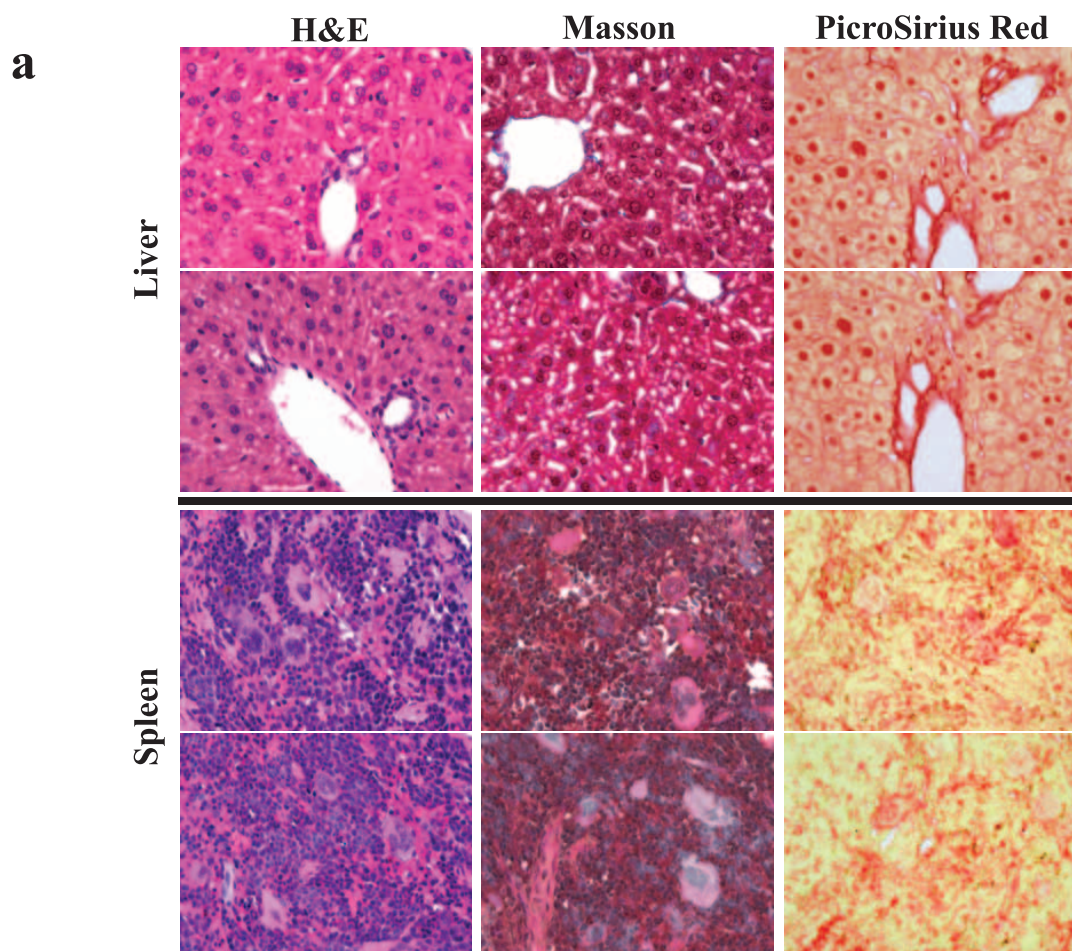
Primary Transplant



Extended Data Figure 9 | Analysis of spleen of primary transplanted mice and bone marrow, spleen and liver of secondary transplanted mice for signs of malignant transformation. a, b, Analysis of spleen of mice 10 months after primary transplantation (from Fig. 3b) of HUVEC-derived rEC-hMPPs as well as bone marrow ($n = 2$), spleen and liver ($n = 2$, also Extended Data Fig. 10a) of mice that were engrafted with secondary transplanted hDMEC-derived rEC-hMPP cells 15 weeks after transplantation (from Fig. 5b) for signs of malignant transformation. The level of fibrosis was determined using Masson and PicroSirius stainings. The architecture of the bone marrow was determined by sequential multi-cross-sectional Wright–Giemsa and haematoxylin and eosin (H&E) staining and compared to age-controlled non-transplanted NSG mice. We did not observe any evidence of fibrosis or alteration of the geometry of the bone marrow, spleen or liver of the transplanted mice. Furthermore, no recipient mouse manifested any anatomical or symptomatic evidence of leukaemias, lymphomas or

myeloproliferative neoplasm (MPN) (that is, lymphadenopathy, splenomegaly/organomegaly, illness or haemorrhage). Circulating hCD45⁺ cells in peripheral blood displayed no evidence of lympho/myeloproliferation or dysplasia. Furthermore, microscopic architecture of bone marrow, spleen and liver was normal and without fibrotic remodelling or aberrant deposition of collagen or desmin. All images were acquired using a colour CCD camera. In primary transplants the scale bar is 200 μ m for low-resolution images in the left columns and 50 μ m for high-resolution images in the right columns. The upper-left image (Giemsa, control) shows a white square in the centre that corresponds to the portion of the image shown at high resolution on the right (the same Giemsa control sample). This rule applies to all images shown (primary transplant). All images in secondary transplant are acquired at an original magnification of $\times 60$. All images are acquired at original magnification of $\times 60$. The top rows of images for each organ are secondary transplants; bottom rows of images for each organ are controls.

Secondary Transplant



Extended Data Figure 10 | Analysis of liver and spleen of secondary transplanted mice for signs of malignant transformation and analyses of rEC-hMPPs for genetic stability. **a**, Analysis of liver and spleen of secondary transplanted mice for signs of malignant transformation. Repeat analysis of spleen and liver of mice that were engrafted with secondary transplanted hDMEC-derived rEC-hMPP cells 15 weeks post-transplantation for signs of malignant transformation (from Fig. 5b). The level of fibrosis was determined by Masson and PicroSirus stainings. The architecture of the bone marrow was determined by sequential multi-cross-sectional haematoxylin and eosin (H&E) staining and compared to age-controlled non-transplanted NSG mice. We did not observe any evidence of fibrosis or alteration of the geometry of the spleen or liver of the transplanted mice. Furthermore, no recipient mouse manifested any anatomical or symptomatic evidence of leukaemias, lymphomas or myeloproliferative neoplasm (MPN) (that is, lymphadenopathy, splenomegaly/organomegaly, illness or haemorrhage). Circulating hCD45⁺ cells in peripheral blood displayed no evidence of lympho/myeloproliferation or dysplasia. Furthermore, microscopic architecture of bone marrow, spleen

and liver was normal and without fibrotic remodelling or aberrant deposition of collagen or desmin. All images are acquired at original magnification of $\times 60$. Top rows of images for each organ are secondary transplants; bottom rows of images for each organ are controls. **b**, Comparative genomic hybridization analysis (CGH) shows that rEC-hMPPs are genetically stable both *in vitro* and *in vivo*. Genomic DNA was extracted from HUVECs, CD45⁺ rEC-hMPPs (35 days post-transduction) or in CD45⁺CD34⁺ rEC-hMPPs sorted from the engrafted NSG bone marrow (24 weeks post-transplantation) and expanded for 72 h *in vitro*. A human tumour sample was used as positive control of chromosome rearrangement. Extracted DNA was digested, labelled by random priming and hybridized to the Agilent 1M CGH arrays. The arrays were scanned in an Agilent DNA microarray scanner and obtained data were visualized using Feature Extraction software (version 10.7; Agilent). No genomic abnormalities were identified in CD45⁺ rEC-hMPPs (or in CD45⁺CD34⁺ rEC-hMPPs) engrafted in NSG bone marrow. Hence, rEC-hMPPs remain genetically stable *in vitro* and *in vivo* and are not transformed.

The cancer glycocalyx mechanically primes integrin-mediated growth and survival

Matthew J. Paszek^{1,2,3,4}, Christopher C. DuFort^{1,2}, Olivier Rossier^{5,6}, Russell Bainer^{1,2}, Janna K. Mouw¹, Kamil Godula^{7,8†}, Jason E. Hudak⁷, Jonathon N. Lakin¹, Amanda C. Wijekoon^{1,2}, Luke Cassereau^{1,2}, Matthew G. Rubashkin^{1,2}, Mark J. Magbanua^{9,10}, Kurt S. Thorn¹¹, Michael W. Davidson¹², Hope S. Rugo^{9,10}, John W. Park^{9,10}, Daniel A. Hammer¹³, Grégory Giannone^{5,6}, Carolyn R. Bertozzi^{7,14,15} & Valerie M. Weaver^{1,2,9,16}

Malignancy is associated with altered expression of glycans and glycoproteins that contribute to the cellular glycocalyx. We constructed a glycoprotein expression signature, which revealed that metastatic tumours upregulate expression of bulky glycoproteins. A computational model predicted that these glycoproteins would influence transmembrane receptor spatial organization and function. We tested this prediction by investigating whether bulky glycoproteins in the glycocalyx promote a tumour phenotype in human cells by increasing integrin adhesion and signalling. Our data revealed that a bulky glycocalyx facilitates integrin clustering by funnelling active integrins into adhesions and altering integrin state by applying tension to matrix-bound integrins, independent of actomyosin contractility. Expression of large tumour-associated glycoproteins in non-transformed mammary cells promoted focal adhesion assembly and facilitated integrin-dependent growth factor signalling to support cell growth and survival. Clinical studies revealed that large glycoproteins are abundantly expressed on circulating tumour cells from patients with advanced disease. Thus, a bulky glycocalyx is a feature of tumour cells that could foster metastasis by mechanically enhancing cell-surface receptor function.

The composition of cell surface glycans and glycoproteins changes markedly and in tandem with cell fate transitions occurring in embryogenesis, tissue development, stem-cell differentiation and diseases such as cancer^{1–3}. Nevertheless, our understanding of the biochemical functions of glycans fails to explain fully why broad changes in glycosylation and glycoprotein expression are critical to cell fate specification and in what ways are they linked to disease. It is currently unclear whether changes in glycan and glycoprotein expression reflect a global and more general mechanism that directs cell and tissue behaviour.

From a materials perspective, glycan and glycoprotein expression dictates the bulk physical properties of the glycocalyx—the exterior cell surface layer across which information flows from the microenvironment to signal transduction pathways originating at the plasma membrane. Although the biophysical functions of the glycocalyx are largely untested, computational models predict that bulky glycoproteins can promote transmembrane receptor organization, including the clustering of integrins at adhesion sites⁴. These models suggest that glycocalyx-mediated integrin clustering would promote the assembly of mature adhesion complexes and collaborate to enhance growth factor signalling⁵—phenotypes that are associated with cancer^{6,7}. We demonstrate that a global modulation of the physical properties of the glycocalyx alters integrin organization and function, and present evidence for how the glycocalyx can be co-opted in malignancy to support tumour cell growth and survival.

Regulation of integrin assembly by bulky glycoproteins

To determine whether glycocalyx bulk contributes to a cancer phenotype, we used gene expression microarray data to relate metastasis to expression of genes for which protein products contribute to the glycocalyx. The likely contribution of gene product to glycocalyx bulk was estimated based on the protein's extracellular domain structure and predicted number of glycosylation sites (Extended Data Fig. 1). Using these estimates we obtained evidence for upregulation of transcripts encoding bulky glycoproteins and some classes of glycosyltransferases, which catalyse the glycosylation of cell surface proteins, in primary tumours of patients with distant metastases relative to those with localized tumour growth ($P = 0.032$ for bulky transmembrane proteins, Kolmogorov–Smirnov test; Fig. 1a and Extended Data Fig. 1).

To understand whether bulky glycoproteins could promote tumour aggression by regulating integrin adhesions, we developed an integrated biochemical and mechanical model that incorporates integrins, the extracellular matrix (ECM), the cell membrane and the glycocalyx (Extended Data Fig. 2). The model revealed that the kinetic rates of integrin–ECM interactions are tightly coupled to the distances between receptor–ligand pairs and, thus, the physical constraints imposed by the glycocalyx. In the presence of bulky glycoproteins, the model predicted that integrin–ECM binding is most favourable at sites of pre-existing adhesive contact, where the membrane and ECM substrate are in closest proximity (Fig. 1b). Elsewhere, bulky glycoproteins sterically restrict efficient

¹Department of Surgery and Center for Bioengineering and Tissue Regeneration, University of California, San Francisco, California 94143, USA. ²Bay Area Physical Sciences-Oncology Program, University of California, Berkeley, California 94720, USA. ³School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, New York 14853, USA. ⁴Laboratory for Atomic and Solid State Physics and Kavli Institute at Cornell for Nanoscale Science, Cornell University, Ithaca, New York 14853, USA. ⁵Interdisciplinary Institute for Neuroscience, University of Bordeaux, UMR 5297, F-33000 Bordeaux, France. ⁶CNRS, Interdisciplinary Institute for Neuroscience, University of Bordeaux, UMR 5297, F-33000 Bordeaux, France. ⁷Department of Chemistry, University of California, Berkeley, California 94720, USA. ⁸The Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. ⁹Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, California 94143, USA. ¹⁰Division of Hematology/Oncology, University of California, San Francisco, California 94143, USA. ¹¹Department of Biochemistry and Biophysics, University of California, San Francisco, California 94158, USA. ¹²National High Magnetic Field Laboratory and Department of Biological Science, The Florida State University, Tallahassee, Florida 32310, USA. ¹³Departments of Chemical and Biomolecular Engineering and Bioengineering, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ¹⁴Department of Molecular Biology, University of California, Berkeley, California 94720, USA. ¹⁵Howard Hughes Medical Institute, University of California, Berkeley, California 94720, USA. ¹⁶Departments of Anatomy and Bioengineering and Therapeutic Sciences and Eli and Edythe Broad Center for Regeneration Medicine and Stem Cell Research, University of California, San Francisco, California 94143, USA. †Present address: Department of Chemistry and Biochemistry, University of California, San Diego, California 92093, USA.

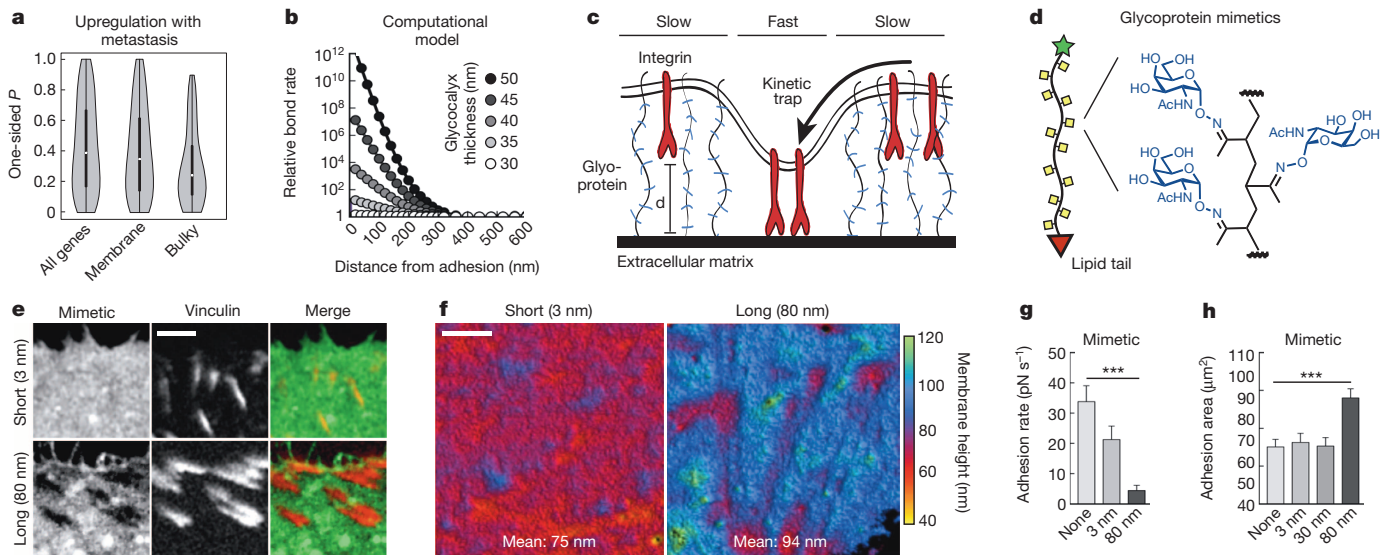


Figure 1 | The cancer glycocalyx drives integrin clustering. **a**, Violin plots showing increased expression of genes encoding bulky transmembrane proteins in primary tumours of patients with distant metastases relative to those with local invasion. White dots and thick black lines indicate the median and interquartile range of the P value distribution of all transcripts within each class: all genes, all membrane proteins (Mem.), and bulky transmembrane proteins (Bulky). **b**, Computed relative rate of integrin–ECM ligand bond formation as a function of distance from a pre-existing adhesion cluster. **c**, Model of proposed glycocalyx-mediated integrin clustering. Shorter distances between integrin–ligand pairs result in faster kinetic rates of binding.

integrin–matrix engagement (Fig. 1b) by increasing the gap between the plasma membrane and ECM. Thus, the model predicted that whereas bulky glycoproteins reduce the overall integrin-binding rate, they enhance, rather than impede, integrin clustering and focal adhesion assembly by generating a physically based kinetic trap (Fig. 1c).

To test experimentally whether bulky glycoproteins could drive integrin clustering and focal adhesion assembly, we generated a series of synthetic mucin glycoprotein mimetics of increasing length that rapidly intercalate into the plasma membrane and project perpendicularly to the cell surface^{8,9}. These glycopolymers consisted of a long-chain polymer backbone, pendant glycan chains that mimic the structures of natural mucin O-glycans, a phospholipid for membrane insertion, and a fluorophore for imaging (Fig. 1d and Extended Data Fig. 3a–c). We found that large glycopolymers with lengths of 80 nm, significantly longer than the reported integrin length of 20 nm¹⁰, are consistently excluded from sites of integrin adhesion on the surface of non-malignant mammary epithelial cells (MECs; Fig. 1e). Shorter polymers with lengths of 3 or 30 nm were not excluded (Fig. 1e). Because the mimetics possessed minimal biochemical interactivity with cell surface proteins (Extended Data Fig. 3d), the data suggest a physical interplay between bulky glycoproteins and integrin receptors.

To determine how the largest polymer mimetics influence the nanoscale spatial features of the cell–ECM interface, we measured the topography of the ventral cell membrane using scanning angle interference microscopy (SAIM), a fluorescence-based microscopy technique that enables imaging with 5–10-nm axial resolution and diffraction-limited (~250 nm) lateral resolution¹¹. Polymers designed to mimic large native glycoproteins (~80 nm) expanded the membrane–ECM gap by 19 nm (Fig. 1f). Consistent with computational predictions, the large glycoprotein mimetics reduced the overall rate of integrin bond formation, but significantly enhanced clustering of integrins into focal adhesions (Fig. 1g, h). Shorter glycoprotein mimetics (3 and 30 nm) did not have an impact on integrin clustering, even when incorporated at higher surface densities (Fig. 1h).

d, Cartoon showing structure of glycoprotein mimetics with lipid insertion domain. **e**, Fluorescence micrographs of MEC adhesion complexes (vinculin–mCherry) and glycomimetics of the indicated length (scale bar, 3 μm). **f**, SAIM images of DiI-labelled ventral plasma membrane topography in MECs incorporated with glycomimetics (scale bar, 2.5 μm). **g**, Rate of integrin–substrate adhesion measured using single cell force spectroscopy in MECs with incorporated glycomimetics. **h**, Quantification of the total adhesion complex area per cell in MECs with incorporated glycomimetics. All results are the mean ± s.e.m. of three separate experiments. Statistical significance is given by * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

We next asked whether cancer-associated glycoproteins could similarly influence the spatial distribution of integrins and the assembly of focal adhesions. On the basis of our large-scale gene expression analysis, we determined that the transmembrane mucin glycoprotein, MUC1, which has a highly glycosylated ectodomain that projects out up to 200 nm from the cell surface¹², was upregulated in metastatic tumours (nominal $P = 0.0028$ via one-sided t -test). In agreement with our analysis, we measured high levels of MUC1 on the surface of several breast cancer cell lines, as well as v-Src and HRAS-transformed MECs (Fig. 2a).

To assess the impact of MUC1 on focal adhesion assembly, we expressed MUC1 on the surface of non-malignant MECs, to levels comparable to those of transformed MECs and breast cancer lines. MUC1 expression induced striking membrane topographical features, which included regions of high curvature, and a significant expansion of the cell membrane–ECM gap (Fig. 2b, c and Extended Data Fig. 4a). Expression of an ectodomain-truncated MUC1 construct did not significantly change the gap compared to control MECs (Fig. 2c and Extended Data Fig. 4a). Our model predicted that the membrane topographies we observed in MUC1-expressing cells would facilitate integrin clustering through the kinetic trap. In agreement with these predictions, expression of full-length MUC1 significantly enhanced the size of adhesion clusters and the total adhesion area per cell (Fig. 2d, e and Extended Data Fig. 5a). The adhesion assembly phenotype did not require the MUC1 cytoplasmic tail, which mediates MUC1's biochemical activity (Fig. 2e)¹³, or direct interactions between MUC1 and fibronectin (Extended Data Fig. 5b). Together, these results are consistent with a physically based mechanism of integrin clustering.

To gain additional insight into the coupled dynamics between integrins and MUC1, we conducted time-lapse imaging of fluorescently labelled MUC1 and the adhesion plaque protein vinculin. We observed that MUC1 and integrin adhesions spatially segregate on the cell surface in a temporally correlated manner (Fig. 2d, Extended Data Fig. 5c and Supplementary Movie 1), suggesting a physical communication between these molecules. Further evidence for a physical interplay between

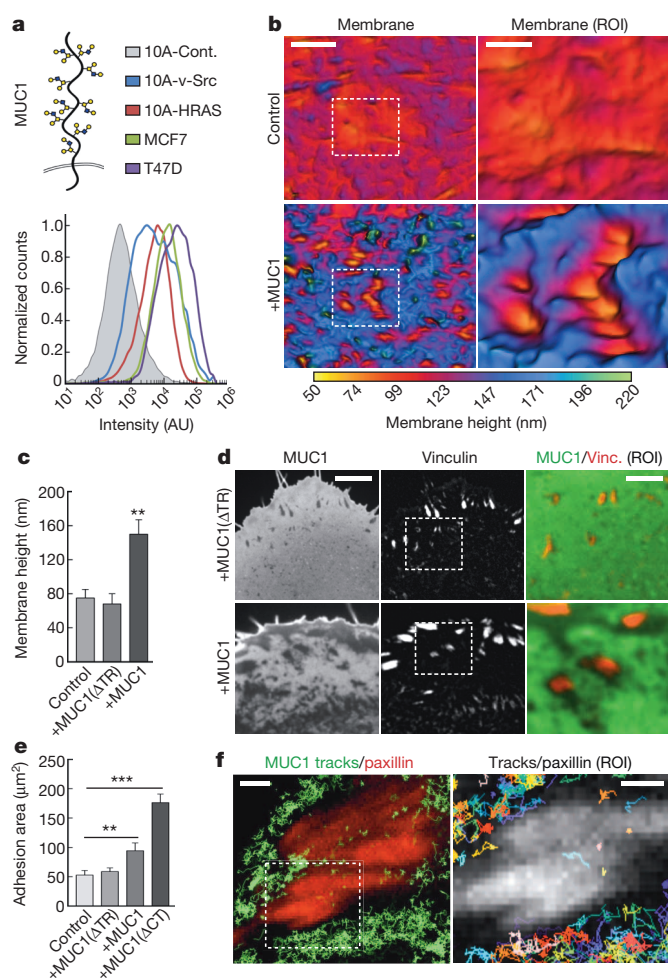


Figure 2 | The bulky cancer-associated glycoprotein MUC1 drives integrin clustering. **a**, Cartoon of MUC1 and quantification of MUC1 cell-surface levels on control (10A-Cont.), transformed (10A-v-Src, 10A-HRAS) and tumour (MCF7, T47D) cells. **b**, Topographical SAIM images of representative mCherry-CAAX-labelled ventral plasma membranes in control and MUC1-GFP-expressing (+MUC1) MECs (Scale bar, 5 μm; region of interest (ROI) scale bar, 2 μm). **c**, Quantification of mean plasma membrane height in control MECs and those ectopically expressing ectodomain-truncated MUC1-GFP (+MUC1(ΔTR)) and wild-type MUC1-GFP (+MUC1). Results are the mean ± s.e.m. of at least 15 cell measurements in duplicate experiments. **d**, Fluorescence micrographs of MUC1(ΔTR) or wild-type MUC1 expressed in MECs and their focal adhesions labelled with vinculin-mCherry (scale bar, 3 μm; ROI scale bar, 1.5 μm). **e**, Quantification of the total adhesion complex area per cell in control non-malignant MECs (control) and those ectopically expressing MUC1(ΔTR), wild-type MUC1, or cytoplasmic-tail-deleted MUC1 (+MUC1(ΔCT)). Results are the mean ± s.e.m. of three separate experiments. **f**, Left panel: trajectories of individual mEOS2-tagged MUC1 proteins recorded at 50 Hz using sptPALM (green) and focal adhesions visualized with paxillin-GFP (red) in MECs (scale bar, 3 μm). Right panel: the ROI from the left panel with individual MUC1 tracks displayed in multiple colours (scale bar, 1 μm). Statistical significance is given by **P* < 0.05; ***P* < 0.01; ****P* < 0.001.

MUC1 and integrins was obtained in mouse embryonic fibroblasts (MEFs) using single-particle tracking photo-activation localization microscopy (sptPALM^{14,15}) to track MUC1 diffusive trajectories. We noted that whereas MUC1 was mobile in the plasma membrane, it rarely crossed into integrin adhesion zones (Fig. 2f and Extended Data Fig. 6a).

We next tested our model's prediction that MUC1 would favour integrin clustering by physically impeding integrin-ECM binding outside of adhesive contacts. We recorded the trajectories of individual β_3 integrin molecules using sptPALM to determine the location and fraction of mobile (confined and freely diffusive) and matrix-bound, immobilized

integrin¹⁵. Analysis of β_3 integrin trajectories after manganese activation in MEFs revealed a significant increase in the total level of immobilized integrin at the plasma membrane, both inside and outside adhesive contacts (Fig. 3a and Extended Data Fig. 6b–e). By contrast, the immobilized β_3 integrin in MEFs expressing high MUC1 was restricted to sites of adhesion (Fig. 3a–c and Extended Data Fig. 6e). These results are consistent with single-cell force spectroscopy measurements, which indicated that MUC1 expression reduces the net rate of integrin-ECM bond formation (Extended Data Fig. 5d). Mucin expression did not have a significant impact on the free diffusion of integrins (Extended Data Fig. 6b–d). Importantly, we observed that integrins frequently diffused across the mucin-adhesive zone boundary and could immobilize rapidly once in the adhesive zone (Fig. 3d, Extended Data Fig. 6f, g and Supplementary Movie 2). Together, our results indicate that large glycoproteins act as physical 'steric' barriers that impede integrin immobilization and thus funnel integrins into adhesive contacts.

Bulky glycoproteins exert force on integrin bonds

Integrins switch between activity states by undergoing a conformational change that is facilitated by tensile force^{16,17}. Given the order of magnitude difference in the size of MUC1 (~200 nm¹²) as compared to integrins (~20 nm¹⁰), and the close proximity of these molecules within the cell-ECM interface, we hypothesized that large glycoproteins, such as MUC1, modify integrin structure and function by applying force to matrix-bound receptors. Abiding by Newton's third law, if large glycoproteins exert a tensile force on integrins, then we should detect a reciprocal strain on the glycoproteins. Consistent with this hypothesis, mucins imaged with SAIM appeared compressed or mechanically bent near integrin adhesive contacts (Fig. 4a and Extended Data Fig. 4a). Furthermore, single-cell force spectroscopy revealed that MECs expressing high levels of exogenous MUC1 required higher compressive force application at the ECM-substrate interface to promote integrin-mediated adhesion when compared to control MECs (Fig. 4b).

To test further whether integrin adhesions strain bulky transmembrane glycoproteins, we generated a genetically encoded construct conceptually similar to a strain gauge, consisting of a cysteine-free cyan and yellow fluorescent protein pair (CFP and YFP) separated by an elastic linker¹⁸, which we inserted into the ectodomain of full-length and truncated MUC1 proteins (Fig. 4c and Extended Data Fig. 4b). Fluorescence resonance energy transfer (FRET) served as the readout of distance between the CFP and YFP pair and, thus, functioned as a reporter of molecular strain. When the full-length reporter was expressed in MECs, we observed high FRET efficiencies in the cell-substrate interface (Fig. 4d, e and Extended Data Fig. 7). FRET efficiency was significantly lower in MECs expressing the ectodomain-truncated construct, indicative of lower molecular strain (Fig. 4d, e). The highest FRET efficiencies correlated spatially with sites of adhesive contact, consistent with integrin adhesions straining bulky transmembrane glycoproteins and glycoproteins exerting a reciprocal restoring force on integrins (Fig. 4d and Extended Data Fig. 7e, f).

We next examined whether the bulky glycoprotein MUC1 could induce conformational changes that would activate integrins independent of the contractile cytoskeleton. We used a bi-functional crosslinker that can specifically link extracellular fibronectin and bound $\alpha_5\beta_1$ integrins that are in a tension-dependent conformation¹⁷. Inhibition of actomyosin contractility, using the myosin inhibitor blebbistatin or the Rho kinase inhibitor Y-27632, abrogated most of the fibronectin crosslinked integrins in MECs expressing empty vector (Fig. 4f and Extended Data Fig. 8a). By contrast, MUC1-expressing MECs formed tensioned bonds with the ECM substrate, even when cells were pre-treated with contractility inhibitors before plating (Fig. 4f and Extended Data Fig. 8a). Of note, the myosin-independent integrin clusters observed in the MUC1-expressing MECs recruited activated cytoplasmic adaptors typically associated with mature adhesion structures and nucleated actin (Extended Data Fig. 8b). These results suggest that large, cancer-associated glycoproteins not only facilitate integrin clustering but also physically alter

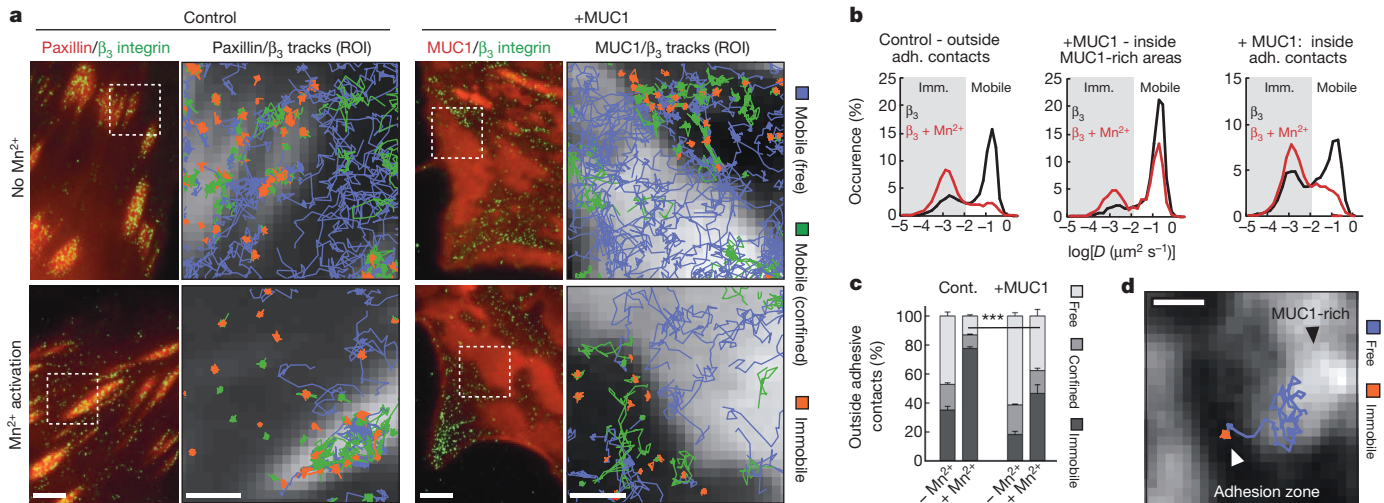


Figure 3 | Bulk glycoproteins spatially regulate immobilization of activated integrins. **a**, Left panels: fluorescence micrographs displaying paxillin-GFP-labelled focal adhesions in control cells or MUC1-rich regions in MUC1-GFP-expressing MEFs, and positions of individual mEOS2-fused β_3 integrins. Cells were treated without or with Mn^{2+} to activate integrins (scale bar, 3 μm). Right panels: magnified area of interest showing fluorescence micrographs of focal adhesions visualized with paxillin-GFP in control MEFs or MUC1 in MUC1-GFP-expressing MEFs, and individual β_3 integrin trajectories recorded with sptPALM. Single molecule trajectories are colour-coded to indicate immobile and mobile (confined and freely diffusive) β_3

integrin state and do so, at least in part, independently of cytoskeletal tension.

Bulky glycoproteins promote growth and survival

Tumour metastasis is a multi-step process that depends on the efficient dissemination of primary cancer cells and their subsequent colonization

integrins (scale bar, 1 μm). **b**, Distribution of β_3 integrin diffusion coefficients recorded before or after Mn^{2+} treatment in control MEFs outside of adhesive contacts (left), MUC1-transfected MEFs inside MUC1-rich areas (middle), and MUC1-transfected MEFs outside MUC1-rich areas, including adhesive contacts (right). **c**, Fraction of immobilized, confined and freely diffusive β_3 integrins outside of adhesive contacts in control MEFs (Ctrl) and MUC1-transfected cells (MUC1) before and after Mn^{2+} treatment. Results are the mean \pm s.e.m. Statistical significance is given by * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. **d**, Fluorescence micrograph of MUC1-GFP and an illustrative single integrin trajectory in MEFs treated with Mn^{2+} (scale bar, 1 μm).

at distant metastatic sites¹⁹. Thus, the ability to survive, particularly within unfavourable microenvironments and under minimally adhesive conditions, is a prerequisite for efficient tumour cell metastasis¹⁹. Given their ability to promote integrin adhesion assembly, we hypothesized that bulky glycoproteins could facilitate metastasis by promoting focal adhesion signalling to enhance tumour cell growth and survival.

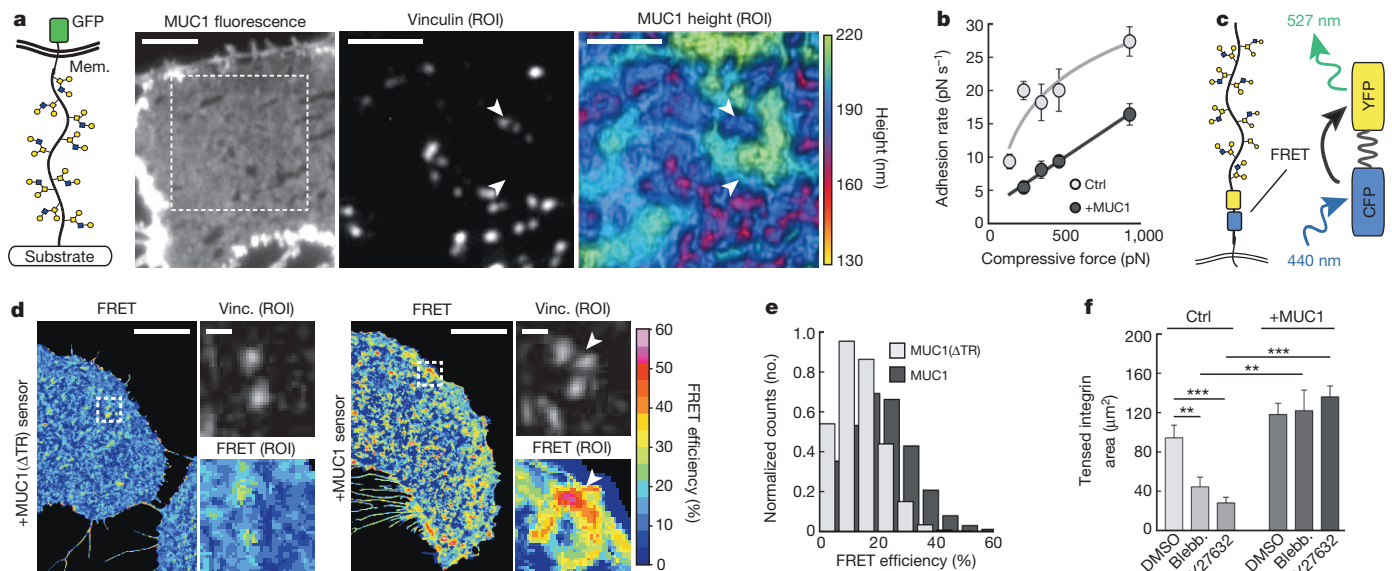


Figure 4 | Integrins are mechanically loaded and re-enforced by bulky glycoproteins. **a**, GFP-fluorescence and topographic SAIM images of MUC1-GFP (scale bars, 3 μm) and the corresponding focal adhesions visualized with vinculin-mCherry. **b**, Adhesion rate versus force of contact between cell and substrate (compressive force) measured with single-cell force spectroscopy for control and MUC1-expressing MECs. Results are the mean \pm s.e.m. of at least 10 cell measurements per point. **c**, Schematic of FRET-based MUC1 compressive strain gauge. **d**, FRET efficiency maps of ectodomain-truncated (+MUC1(ΔTR) sensor) and wild-type (+MUC1 sensor) strain gauges measured at the ventral cell surface of MECs and the

corresponding vinculin-mCherry-labelled focal adhesions (scale bar, 8 μm ; ROI scale bar, 1 μm). **e**, Histogram of observed FRET efficiencies of wild-type MUC1 and MUC1(ΔTR) strain gauges. **f**, Quantification of fibronectin-crosslinked α_5 integrin in control and MUC1-expressing normal MECs treated with solvent alone (DMSO), myosin-II inhibitor (blebbistatin; 50 μM), or Rho kinase inhibitor (Y-27632; 10 μM) for 1 h followed by detergent-extraction to reveal the fibronectin bound integrin that is under mechanical tension. Results are the mean \pm s.e.m. of three separate experiments. Statistical significance is given by * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

Consistent with this notion, analysis of human data sets revealed that patients with aggressive breast cancers that presented with circulating tumour cells (CTCs) express disproportionately high amounts of bulky glycoproteins and have altered glycosyltransferase expression profiles (Fig. 5a and Extended Data Fig. 1d, e). Furthermore, analysis of genes expressed within CTCs isolated from a cohort of breast cancer patients with metastatic disease confirmed that several predicted bulky glycoproteins could be detected in these patient samples (Fig. 5b).

We next examined whether a bulky glycocalyx could promote growth and survival of non-malignant MECs. Using our glycoprotein mimetics, we observed that untreated MECs or MECs incorporated with short (3 nm) or medium (30 nm) length mimetics were not viable 24–48 h after they were plated on highly compliant hydrogel substrates that mimic the stiffness of soft sites of colonization, like lung or brain (Young's modulus, $E = 140$ Pa; Fig. 5c). By contrast, MECs incorporated with long glycoprotein mimetics (80 nm) remained viable (Fig. 5c). Analysis of gene expression profiles and immunofluorescence analysis of freshly isolated CTCs in our human metastatic breast cancer cohort revealed that MUC1 could be detected in most of the samples examined (Fig. 5b). Similar to results with the synthetic mimetics, we observed that ectopic expression of either full-length or a tailless, signalling-defective MUC1 in non-malignant MECs permitted their growth and survival even when plated as single cells on compliant hydrogels (Fig. 5d and Extended Data Fig. 9a).

We noted that the CTCs in our cohort also expressed high levels of CD44, a receptor that binds and retains bulky hyaluronic acid (HA) glycan structures on the cell surface (Extended Data Fig. 10a)²⁰. Similar to our observations with MUC1 and bulky glycoprotein mimetics, we observed that HA and integrins exhibit an anti-correlated spatial

distribution on the surface of transformed MECs (Extended Data Fig. 10b). Inhibition of HA synthesis or HA cell-surface retention significantly reduced the growth of transformed MECs on compliant hydrogels, raising the possibility that bulky cell-surface constituents, in addition to MUC1, could similarly promote tumour aggression (Extended Data Fig. 10b). However, unlike the experiments with tailless MUC1 or the glycoprotein mimetics, which lack signalling capability, we cannot exclude that HA-induced growth and survival phenotypes are not also, at least in part, induced through HA's direct biochemical signalling activity^{20,21}.

We next addressed whether a bulky glycocalyx promotes MEC growth and survival by regulating focal adhesion assembly and crosstalk with growth factor signalling pathways^{5,7}. We found that pharmacological inhibition of kinases linked to growth factor signalling, including phosphoinositide 3-kinase (PI(3)K), mitogen-activated kinase, and Src kinase, each independently inhibited the growth and survival of MUC1-expressing MECs on highly compliant substrates (Fig. 5e). We also noted that the MUC1 growth and survival phenotype requires integrin engagement and integrin signalling through focal adhesion kinase (FAK), which mediates crosstalk between integrin and growth factor signalling pathways (Fig. 5f and Extended Data Fig. 9b)^{5,6}. Non-malignant MECs expressing the MUC1 ectodomain, but not control MECs, assembled distinct focal adhesion structures with activated Y397-phosphorylated FAK on compliant substrates (Extended Data Fig. 8c). Furthermore, MECs expressing wild-type or tailless, signalling defective MUC1, and plated on the compliant substrates showed enhanced Y118-phosphorylated paxillin, ERK and AKT activation in response to epidermal growth factor stimulation (Fig. 5g and Extended Data Fig. 8d). This response was attenuated by FAK inhibition (Fig. 5g, h and Extended Data Fig. 8d).

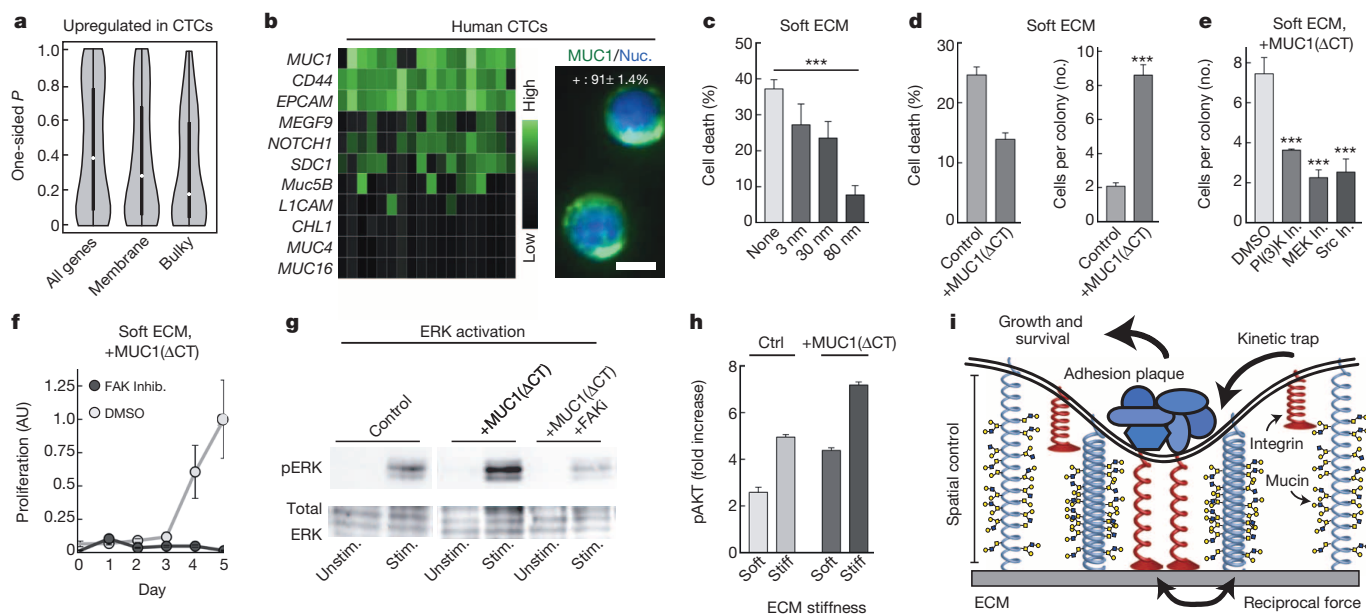


Figure 5 | Bulky glycoproteins promote cell survival and are expressed in CTCs. **a**, Violin plots showing that genes encoding bulky transmembrane proteins are more highly expressed in primary human tumours in patients with circulating tumour cells (CTCs). White dots and thick black lines indicate the median and interquartile range of the P -value distribution of transcripts of all cellular genes (all genes), all transmembrane proteins (membrane), and bulky transmembrane proteins (bulky). **b**, Heat map quantifying gene expression of bulky glycoproteins in CTCs isolated from 18 breast cancer patients (x axis; left), and representative immunofluorescence micrograph of MUC1 detected on human patient CTCs (right; scale bar, 5 μ m). Quantification of the percentage of CTCs with detectable MUC1 is shown. **c**, Cell death in control non-malignant MECs and those with incorporated glycomimetics quantified 24 h after plating on a soft (140 Pa) fibronectin-conjugated hydrogel substrate. **d**, Cell death (left graph) and growth (right graph) of control MECs and those expressing cytoplasmic-tail-deleted MUC1 (+MUC1(Δ CT))

quantified 48 h after plating on a soft hydrogel. **e**, Quantification of the number of vehicle (DMSO), PI(3)K inhibitor, MEK inhibitor, or Src inhibitor-treated control and MUC1(Δ CT)-expressing MECs per colony 48 h after plating on a soft hydrogel. **f**, Proliferation of solvent (DMSO) or FAK-inhibitor-treated MUC1(Δ CT)-expressing MECs quantified at the indicated day after plating on soft hydrogels. **g**, Representative western blots showing phosphorylated and total ERK in control and MUC1(Δ CT)-expressing MECs plated on soft hydrogels unstimulated or stimulated with EGF. Cells were treated with solvent (control, +MUC1(Δ CT)) or FAK inhibitor (+MUC1(Δ CT) + FAKi) before stimulation. **h**, Bar graphs showing quantification of immunoblots probed for activated AKT in control and MUC1(Δ CT)-expressing non-malignant MECs 24 h after plating on soft versus stiff hydrogels. **i**, Model summarizing biophysical regulation of integrin-dependent growth and survival by bulky glycoproteins. In all bar graphs, results are the mean \pm s.e.m. of at least 2–3 separate experiments ($^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$).

Together, these findings indicate that a bulky glycocalyx can promote tumour aggression by enhancing integrin-dependent growth and survival (Fig. 5i).

Discussion

We present evidence to support a new paradigm for the biological function of cell surface glycans and glycoproteins. Independent of, and in addition to, their biochemical properties, we demonstrate how bulky constituents of the glycocalyx can physically influence receptor organization and activity. Although the current investigation focuses on integrins, a bulky glycocalyx could, in principle, regulate any transmembrane receptor that interacts with a tethered ligand. Candidate systems include neurological and immunological synapses²², cell–cell adhesions²³, and juxtacrine signalling complexes composed of receptors, like ephrin²⁴. Membrane topographical features imprinted by large glycoproteins could also directly influence plasma membrane lipid organization, protein sorting and endocytosis^{25,26}. The diversity of these processes suggests that the physiological relevance of the glycocalyx may be broad. For example, bulky glycoproteins and glycan structures, such as neuroligins, neurexins and polysialic acid, have a crucial role in neuronal development, maintenance and plasticity^{27,28}. Thus, it is plausible that the glycocalyx has a prominent role in orchestrating multiple biological processes occurring at the plasma membrane.

Our observations provide a tractable explanation for why large glycan structures and glycoproteins, like HA and mucins, as well as regulatory enzymes, are so frequently elevated in many solid tumours^{13,20}. Indeed, the growth and survival advantages afforded by these molecules may preferentially select for cancer cells with a prominent glycocalyx and favour tumour cell dissemination and metastasis. Mechanical perturbations to cell and tissue structure have a causal role in tumour development and progression^{29,30}, and we now implicate the glycocalyx's importance in the metastatic mechano-phenotype. Our results suggest that the glycocalyx and its molecular constituents are attractive targets for therapeutic interventions aimed at normalizing transmembrane receptor signalling.

METHODS SUMMARY

Complete descriptions of the bioinformatics pipeline, computational model and expression constructs are presented in Supplementary Notes 1, 2 and 3, respectively. Compliant hydrogels were fabricated from soft polyacrylamide ($E = 140$ Pa) functionalized with fibronectin³¹ and plated with single cells for all hydrogel experiments. FRET measurements were conducted in living cells on a spinning disk confocal (photobleaching FRET) or confocal (lifetime imaging) microscope³² (see also Supplementary Note 4). SptPALM¹⁵, SAIM¹¹, single cell force spectroscopy³³, integrin crosslinking¹⁷, and fibronectin fibrillogenesis³⁴ measurements and assays were conducted as previously described. Glycoprotein mimetics were synthesized and characterized as described in Supplementary Note 6 and subsequently incubated with suspended cells ($2 \mu\text{M}$ for 1 h) to incorporate onto the cell surface immediately before experimentation. For gene expression analysis of CTCs, 20 pools of CTCs were isolated from the blood of 18 metastatic breast cancer patients and quantified by qPCR³⁵. Immunofluorescence of CTCs was conducted on samples isolated from three patients³⁵.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 3 August 2013; accepted 23 May 2014.

Published online 25 June 2014.

1. Lancot, P. M., Gage, F. H. & Varki, A. P. The glycans of stem cells. *Curr. Opin. Chem. Biol.* **11**, 373–380 (2007).
2. Haltiwanger, R. S. & Lowe, J. B. Role of glycosylation in development. *Annu. Rev. Biochem.* **73**, 491–537 (2004).
3. Varki, A., Kannagi, R. & Toole, B. P. in *Essentials Glycobiol.* (Varki, A. et al.) <http://www.ncbi.nlm.nih.gov/books/NBK1963/> (Cold Spring Harbor Laboratory Press, 2009).
4. Paszek, M. J., Boettiger, D., Weaver, V. M. & Hammer, D. A. Integrin clustering is driven by mechanical resistance from the glycocalyx and the substrate. *PLOS Comput. Biol.* **5**, e1000604 (2009).

5. Walker, J. L., Fournier, A. K. & Assoian, R. K. Regulation of growth factor signaling and cell cycle progression by cell adhesion and adhesion-dependent changes in cellular tension. *Cytokine Growth Factor Rev.* **16**, 395–405 (2005).
6. Desgrosellier, J. S. & Cheres, D. A. Integrins in cancer: biological implications and therapeutic opportunities. *Nature Rev. Cancer* **10**, 9–22 (2010).
7. Shibue, T. & Weinberg, R. A. Integrin $\beta 1$ -focal adhesion kinase signaling directs the proliferation of metastatic cancer cells disseminated in the lungs. *Proc. Natl Acad. Sci. USA* **106**, 10290–10295 (2009).
8. Godula, K. et al. Control of the molecular orientation of membrane-anchored biomimetic glycopolymers. *J. Am. Chem. Soc.* **131**, 10263–10268 (2009).
9. Godula, K., Rabuka, D., Nam, K. T. & Bertozzi, C. R. Synthesis and microcontact printing of dual end-functionalized mucin-like glycopolymers for microarray applications. *Angew. Chem. Int. Edn Engl.* **48**, 4973–4976 (2009).
10. Eng, E. T., Smaghe, B. J., Walz, T. & Springer, T. A. Intact $\alpha \text{IIb} \beta 3$ integrin is extended after activation as measured by solution X-ray scattering and electron microscopy. *J. Biol. Chem.* **286**, 35218–35226 (2011).
11. Paszek, M. J. et al. Scanning angle interference microscopy reveals cell dynamics at the nanoscale. *Nature Methods* **9**, 825–827 (2012).
12. Hattrop, C. L. & Gendler, S. J. Structure and function of the cell surface (tethered) mucins. *Annu. Rev. Physiol.* **70**, 431–457 (2008).
13. Kufe, D. W. Mucins in cancer: function, prognosis and therapy. *Nature Rev. Cancer* **9**, 874–885 (2009).
14. Manley, S. et al. High-density mapping of single-molecule trajectories with photoactivated localization microscopy. *Nature Methods* **5**, 155–157 (2008).
15. Rossier, O. et al. Integrins $\beta 1$ and $\beta 3$ exhibit distinct dynamic nanoscale organizations inside focal adhesions. *Nature Cell Biol.* **14**, 1057–1067 (2012).
16. Chen, W., Lou, J., Evans, E. A. & Zhu, C. Observing force-regulated conformational changes and ligand dissociation from a single integrin on cells. *J. Cell Biol.* **199**, 497–512 (2012).
17. Friedland, J. C., Lee, M. H. & Boettiger, D. Mechanically activated integrin switch controls $\alpha 5 \beta 1$ function. *Science* **323**, 642–644 (2009).
18. Grashoff, C. et al. Measuring mechanical tension across vinculin reveals regulation of focal adhesion dynamics. *Nature* **466**, 263–266 (2010).
19. Nguyen, D. X., Bos, P. D. & Massagué, J. Metastasis: from dissemination to organ-specific colonization. *Nature Rev. Cancer* **9**, 274–284 (2009).
20. Toole, B. P. Hyaluronan: from extracellular glue to pericellular cue. *Nature Rev. Cancer* **4**, 528–539 (2004).
21. Bono, P., Rubin, K., Higgins, J. M. & Hynes, R. O. Layilin, a novel integral membrane protein, is a hyaluronan receptor. *Mol. Biol. Cell* **12**, 891–900 (2001).
22. Dustin, M. L. Signaling at neuro/immune synapses. *J. Clin. Invest.* **122**, 1149–1155 (2012).
23. Coombs, D., Dembo, M., Wofsy, C. & Goldstein, B. Equilibrium thermodynamics of cell–cell adhesion mediated by multiple ligand–receptor pairs. *Biophys. J.* **86**, 1408–1423 (2004).
24. Salaita, K. et al. Restriction of receptor movement alters cellular response: physical force sensing by EphA2. *Science* **327**, 1380–1385 (2010).
25. Groves, J. T. Bending mechanics and molecular organization in biological membranes. *Annu. Rev. Phys. Chem.* **58**, 697–717 (2007).
26. Lundmark, R. & Carlsson, S. R. Driving membrane curvature in clathrin-dependent and clathrin-independent endocytosis. *Semin. Cell Dev. Biol.* **21**, 363–370 (2010).
27. Comoletti, D. et al. Synaptic arrangement of the neuroligin/ β -neurexin complex revealed by X-ray and neutron scattering. *Structure* **15**, 693–705 (2007).
28. Hildebrandt, H., Mühlenhoff, M., Weinhold, B. & Gerardy-Schahn, R. Dissecting polysialic acid and NCAM functions in brain development. *J. Neurochem.* **103** (Suppl. 1), 56–64 (2007).
29. Levental, K. R. et al. Matrix crosslinking forces tumor progression by enhancing integrin signaling. *Cell* **139**, 891–906 (2009).
30. Paszek, M. J. et al. Tensional homeostasis and the malignant phenotype. *Cancer Cell* **8**, 241–254 (2005).
31. Lakin, J. N., Chin, A. R. & Weaver, V. M. Exploring the link between human embryonic stem cell organization and fate using tension-calibrated extracellular matrix functionalized polyacrylamide gels. *Methods Mol. Biol.* **916**, 317–350 (2012).
32. Bruns, N., Pustelný, K., Bergeron, L. M., Whitehead, T. A. & Clark, D. S. Mechanical nanosensor based on FRET within a thermosome: damage-reporting polymeric materials. *Angew. Chem. Int. Ed.* **48**, 5666–5669 (2009).
33. Friedrichs, J., Helenius, J. & Müller, D. J. Quantifying cellular adhesion to extracellular matrix components by single-cell force spectroscopy. *Nature Protocols* **5**, 1353–1361 (2010).
34. Pankov, R. & Momchilova, A. Fluorescent labeling techniques for investigation of fibronectin fibrillogenesis (labeling fibronectin fibrillogenesis). *Methods Mol. Biol.* **522**, 261–274 (2009).
35. Magbanua, M. J. M. et al. Genomic profiling of isolated circulating tumor cells from metastatic breast cancer patients. *Cancer Res.* **73**, 30–40 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Gendler, J. Goedhart and M. McMahon for cDNAs, as indicated in the Methods section. We thank A. Walker for bioinformatics support, L. Hauran for assistance in CTC analysis, H. Aaron for assistance with FLIM, J. B. Sibarta and M. Lagardère for support in sptPALM analysis, B. Hoffman in design of the FRET sensor, and T. Wittmann in assistance with pbFRET measurements. Image acquisition was partly performed at the Nikon Imaging Center and Biological Imaging Development Center at UCSF and the Berkeley Molecular Imaging Center. This work was supported by the Kavli Institute and UCSF Program for Biomedical Breakthrough postdoctoral fellowships to M.J.P.; DoD NDEG Fellowship to M.G.R.; NIH GM59907 to C.R.B.; NIH Pathway to Independence Award K99 EB013446-02 to K.G.; French

Ministry of Research, CNRS, ANR grant Nanomotility, INSERM, Fondation ARC pour la Recherche sur le Cancer, France Biolmaging ANR-10-INBS-04-01, and Conseil Régional Aquitaine to O.R. and G.G.; NIH AI082292-03A1 to D.A.H.; The Breast Cancer Research Foundation to M.J.M., H.S.R. and J.W.P.; NIH 2R01GM059907-13 to C.R.B. and V.M.W.; and BCRP DOD Era of Hope Scholar Expansion grant BC122990, and NIH NCI grants U54CA163155-01, U54CA143836-01, 1U01 ES019458-01, and CA138818-01A1 to V.M.W.

Author Contributions V.M.W. and M.J.P. conceived the project. V.M.W., M.J.P. and C.R.B. provided project management. M.J.P., A.C.W., M.W.D. and J.N.L. designed and constructed expression constructs. C.C.D. and M.J.P. conducted single cell force spectroscopy measurements. M.J.P. and K.S.T. designed and implemented the interference microscope. M.J.P., L.C. and M.G.R. performed fluorescence, FRET and interferometric imaging and M.J.P. wrote the accompanying analysis software. O.R. and

G.G. conducted sptPALM experiments and analysed the results. V.M.W., R.B. and M.J.P. designed the bioinformatics pipeline and R.B. implemented the pipeline and performed the large-scale gene expression analysis. J.K.M., A.C.W. and M.J.P. fabricated and conducted experiments on compliant hydrogels. M.J.M., H.S.R. and J.W.P. isolated CTCs and measured CTC gene and protein expression. K.G., J.E.H. and C.R.B. designed, synthesized and characterized the glycoprotein mimetics. M.J.P. and D.A.H. constructed and implemented the computational model. M.J.P. and V.M.W. wrote the manuscript with input from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to V.M.W. (Valerie.Weaver@ucsfmedctr.org).

METHODS

Bioinformatics. To estimate protein-level contributions to extracellular membrane bulkiness, we used TMHMM to identify extracellular domains within each isoform sequence (RefSeq v47) and counted the number of putative extracellular glycosylation sites predicted by NetOGlyc 3.1 and search of N-glycosylation motifs. Gene-wise enrichment of mRNA upregulation among bulky proteins in clinical data (GEO accessions GSE12276 and GSE31364) was tested by permuting *P* values quantifying evidence for upregulation in the appropriate samples. Variance in mRNA upregulation explained by membrane bulkiness was estimated by regressing the negative log-transformed *P* values on the square root of the combined N- and O-glycosylation sites and comparing the residuals with the intercept model. Additional details of the analysis and models are provided in Supplementary Note 1.

Computational model. A mechanical model of the cell–ECM interface was constructed as described previously⁴. A summary of the model is described in Supplementary Note 2 and parameters are detailed in Supplementary Table 1.

Antibodies and reagents. Antibodies used include: mouse monoclonal antibody (mAb) vinculin (MAB674; Millipore), mouse mAb talin (8d4; Sigma), rat mAb β_1 -integrin (AIIBII), rabbit mAb paxillin (Y113; Abcam), rabbit mAb FAK pY397 (141-9; Invitrogen), rabbit polyclonal antibody (pAb) α_5 -integrin (AB1928; Millipore), mouse mAb MUC1 (HMPV; BD Pharmingen), hamster mAb MUC1 (CT2; Thermo Scientific), rabbit mAb Src Family pY416 (D49G4; Cell Signaling), mouse mAb FAK (77; BD Transduction Laboratories), rabbit pAb paxillin pY118 (2541; Cell Signaling), rabbit mAb pan-AKT (C67E7; Cell Signaling), rabbit pAb AKT pS473 (9271; Cell Signaling); rabbit mAb ERK1/2 pT202/pT204 (197G2; Cell Signaling); rabbit pAb ERK1/2 (9102; Cell Signaling); rabbit mAb Gapdh (14C10; Cell Signaling); Alexa 488 and Alexa 568 conjugated goat anti-mouse and anti-rabbit mAbs (Invitrogen); FITC conjugated anti-hamster mAbs; Cy5-conjugated goat anti-mouse and rabbit mAbs (Jackson); and HRP conjugated anti-rabbit and anti-mouse mAbs. Chemical inhibitors used in these studies include ROCK inhibitor Y-27632 (Cayman Chemical), myosin-II inhibitor (-)-blebbistatin (Cayman Chemical), FAK inhibitor FAK inhibitor 14 (Tocris), MEK inhibitor U0126 (Cell Signaling), PI(3)K inhibitor Wortmannin (Cell Signaling), Src inhibitor Src I1 (Sigma), and DiI (Molecular Probes).

Cell culture conditions. All cells were maintained at 37 °C and 5% CO₂. MCF10A human MECs (ATCC) were cultured in DMEM F12 (Invitrogen) supplemented with 5% donor horse serum (Invitrogen), 20 ng ml⁻¹ epidermal growth factor (Peprotech), 10 μ g ml⁻¹ insulin (Sigma), 0.5 μ g ml⁻¹ hydrocortisone (Sigma), 0.1 μ g ml⁻¹ cholera toxin (Sigma), and 100 units ml⁻¹ penicillin/streptomycin. MCF7 and T47D breast tumour lines (ATCC) were grown in DMEM supplemented with 10% fetal bovine serum (Hyclone) and 100 units ml⁻¹ penicillin/streptomycin. 293T cells (ATCC) were maintained in DMEM supplemented with 10% donor horse serum, 2 mM L-glutamine, and penicillin/streptomycin. Mouse embryonic fibroblasts (MEFs) were cultured in DMEM with 10% fetal bovine serum. Cell lines were tested routinely for mycoplasma contamination. For transient gene expression in MECs, constructs in pcDNA3.1 or Clontech-style vectors were nucleofected with Kit V (Lonza) using program T-024 24 h before experimentation. Transient transfection in MEFs was conducted 48 h before experimentation using Fugene 6 (Roche) or nucleofection. For stable cell lines harbouring tetracycline inducible transgenes, expression was induced with 0.2 ng ml⁻¹ doxycycline 24 h before experimentation. The conditional v-Src oestrogen receptor fusion (v-Src-ER) was activated with 1 μ M 4-hydroxytamoxifen 48 h before experimentation to achieve transformation. For pERK, pY118-paxillin, and pAKT studies, cells were plated on fibronectin-conjugated polyacrylamide hydrogels, serum-starved overnight, and stimulated with 20 ng ml⁻¹ EGF before collecting protein lysates. Data are reported as the fold increase of phosphorylated protein relative to total protein, following EGF stimulation.

Preparation of cellular substrates. Glass and silicon substrates were prepared by glutaraldehyde activation followed by conjugation with 10 μ g ml⁻¹ (glass) or 20 μ g ml⁻¹ (silicon) fibronectin as described¹¹. Compliant polyacrylamide hydrogel substrates (soft: 2.5% acrylamide, 0.03% Bis-acrylamide; stiff: 10% acrylamide, 0.5% Bis-acrylamide) were prepared as previously described with one modification: functionalization with succinimidyl ester was with 0.01% N6, 0.01% Bis-acrylamide, 0.025% Irgacure 2959, and 0.002% Di(trimethylolpropane) tetraacrylate (Sigma)³¹. Following functionalization with succinimidyl ester, hydrogels were conjugated overnight with 20 μ g ml⁻¹ fibronectin at 4 °C and rinsed twice with PBS and DMEM before cell plating.

Generation of expression constructs. A description of cDNA constructs and their construction is provided in Supplementary Note 3.

Generation of stable cell lines. Stable transgene expression was achieved through retroviral or lentiviral transduction as previously described^{11,30}.

Flow cytometry and cell sorting. Cell surface MUC1 was labelled directly with FITC-conjugated mAb MUC1 (clone HMPV). Cytometry and sorting were conducted on a FACSAria II (BD Biosciences).

Immunofluorescence and imaging. Cells were fixed and labelled as previously described and imaged at random on a Zeiss LSM 510 microscope system with a 100X Plan Apochromat NA 1.4 objective and 488 nm Argon, 543 nm HeNe, and 633 nm HeNe excitation lines³⁰.

Live epithelial cell imaging and FRET. Normal growth media was exchanged for a similar formulation lacking phenol red and supplemented with 15 mM HEPES buffer, pH 7.4. Cells were imaged on a Ti-E Perfect Focus System (Nikon) equipped with a CSU-X1 spinning disk confocal unit; 454 nm, 488 nm, 515 nm and 561 nm lasers; an Apo TIRF 60X NA 1.49 objective; electronic shutters; a charged-coupled device camera (Clara; Andor) and controlled by NIS-Elements software (Nikon).

For measurement of FRET efficiency, the acceptor photobleaching method pbFRET was implemented with live cells on the spinning disk confocal. Cyan fluorescent protein (CFP) was first imaged with 454 nm excitation and a 480/20 emission filter, yellow fluorescent protein (YFP) was subsequently bleached using a 100 mW 515 nm laser for 10 s, and CFP was imaged again following bleaching of YFP. Microscope Z-focus was maintained during image acquisition using the Perfect Focus System. Background images were constructed by imaging 10 unique cell-free regions on the coverslip and averaging the intensity at each pixel. The FRET efficiency was calculated on a pixel-by-pixel basis according to:

$$\text{FRET efficiency (\%)} = \left[1 - \frac{I_{\text{pre}} - B_{\text{pre}}}{I_{\text{post}} - B_{\text{post}}} \right] 100\%$$

where I_{pre} is the CFP intensity before bleaching YFP, B_{pre} is the CFP-channel background intensity before bleaching YFP, I_{post} is the CFP intensity after bleaching YFP, and B_{post} is the CFP-channel background intensity after bleaching YFP. Appropriate controls were implemented to account for inadvertent CFP photobleaching, incomplete YFP photobleaching, and intermolecular FRET (see Supplementary Note 4).

Time-domain fluorescence lifetime imaging microscopy (FLIM) for additional FRET sensor characterization was implemented with an inverted Zeiss LSM510 Axiovert 200M microscope with a Plan NeoFLUAR 40X/1.3 NA DIC oil-immersion objective lens, equipped with a TCSPC controller (SPC-830 card; Becker & Hickl, Berlin, Germany) as described previously³². CFP was excited with 440 nm light generated by frequency doubling of 880 nm pulses from a mode-locked Ti:sapphire laser (Mai-Tai, Spectra-Physics, 120–150 fs pulse width, 80 MHz repetition rate, and Frequency Doubler and Pulse Selector, Spectra-Physics, Model 3980). The emission light was passed through a NFT 440 beamsplitter, directed to the fibre-out port of the confocal scan-head, filtered with a 480BP40 filter (Chroma Technology, Rockingham, VT) and detected by a PMC-100 photomultiplier (Becker & Hickl). The pinhole was set to give an optical slice of <4.0 μ m. Images of 386 \times 386 pixels were averaged over <120 s. Data analysis to produce an intensity image and a FLIM image was done off-line using the pixel-based fitting software SPCImage (Becker & Hickl), assuming double exponential decay during the first 8.5 ns of the 12.5 ns interval between laser pulses. Images were scaled to 256 \times 256 pixels and no binning was used. Lifetime distributions were calculated for a masked portion of the FLIM image, generated with a triangle algorithm threshold of the photo count intensity image.

Scanning angle interference microscopy. Cells were plated overnight on reflective silicon substrates, fixed or roofed to remove the dorsal membrane (for MUC1-GFP imaging) and then fixed, and imaged randomly as previously described, scanning the incident angle of excitation light from 0° to 42° with a one-degree sampling rate¹¹. Z-positions were localized with custom algorithms previously described and available on request¹¹.

Single particle tracking photo-activation localization microscopy (sptPALM). sptPALM experiments were performed and analysed as previously described¹⁵. Briefly, live MEFs were imaged at 37 °C in a Ludin chamber on a Ti Perfect Focus System equipped with a Plan Apo 100X NA 1.45 objective, and an electron multiplying charge-coupled device (Evolve; Photometrics). For photo-activation localization, cells expressing mEOS2-tagged constructs were activated using a 405 nm laser (Omicron) and the photo-activated fluorophores were excited simultaneously with a 561 nm laser (Cobolt Jive). The powers of the activation and excitation lasers were adjusted to keep the number of activated molecules constant and well separated. GFP fusions of paxillin or MUC1 were imaged in between each sptPALM sequence by imaging the GFP signal above the unconverted mEOS2 background. The acquisition was driven by Metamorph software (Molecular Devices) in streaming mode at 50 Hz. For tracking, single-molecules were localized and tracked over time using a combination of wavelet segmentation and simulated annealing algorithms. Trajectories lasting at least 20 frames were selected for further quantification, including calculation of immobile, confined and free-diffusing fraction (see Supplementary Note 5)¹⁵.

Preparation of glycopolymer-coated cell surfaces. Mucin mimetic glycopolymers with lipid insertion domains were synthesized and characterized as described in Supplementary Note 6. For incorporation into the plasma membrane, cells were

suspended in DMEM and incubated with 2 μ M glycopolymer for 1 h. Cells were pelleted by centrifugation and re-suspended in growth media to remove unincorporated polymer.

Quantification of adhesion complexes. Images of adhesions in fixed, immuno-labelled cells or cells expressing paxillin-mCherry were randomly acquired, smoothed with a median filter, and background subtracted (12 pixel diameter) in ImageJ. Adhesion sizes and the number of adhesions per cell were subsequently quantified in ImageJ with the 'Analyze Particles' tool.

Integrin crosslinking assay. Cells were incubated in suspension with inhibitor (Y-27632 or Blebbistatin) or control solvent for 1 h before plating on glass substrates. Integrin was crosslinked to fibronectin with 1 mM 3,3'-dithiobis(sulfosuccinimidylpropionate) (Pierce Chemical) and cells were extracted with SDS buffer as previously described¹⁷. Crosslinked α 5 integrin was immuno-labelled and imaged at random with a Plan Apo VC 60X objective on a Nikon TE2000 epi-fluorescence microscope equipped with a charged-coupled device camera (HQ2; Photometrics).

Single cell force spectroscopy. Measurements were performed on an Asylum MFP-3D-BIO atomic force microscope as previously described³³. Briefly, cells were attached to a streptavidin-coated, tipless cantilever using biotinylated jacalin (MUC1-expressing cells) or concanavalin A (all other cells) and pressed against the adhesive substrate with a calibrated force and duration before measuring the force required to detach the cell from the substrate. All measurements were conducted on fibronectin- or BSA-coated glass slides at room temperature. The relative rate of adhesion was calculated as the slope of a linear fit of cellular detachment force against contact time.

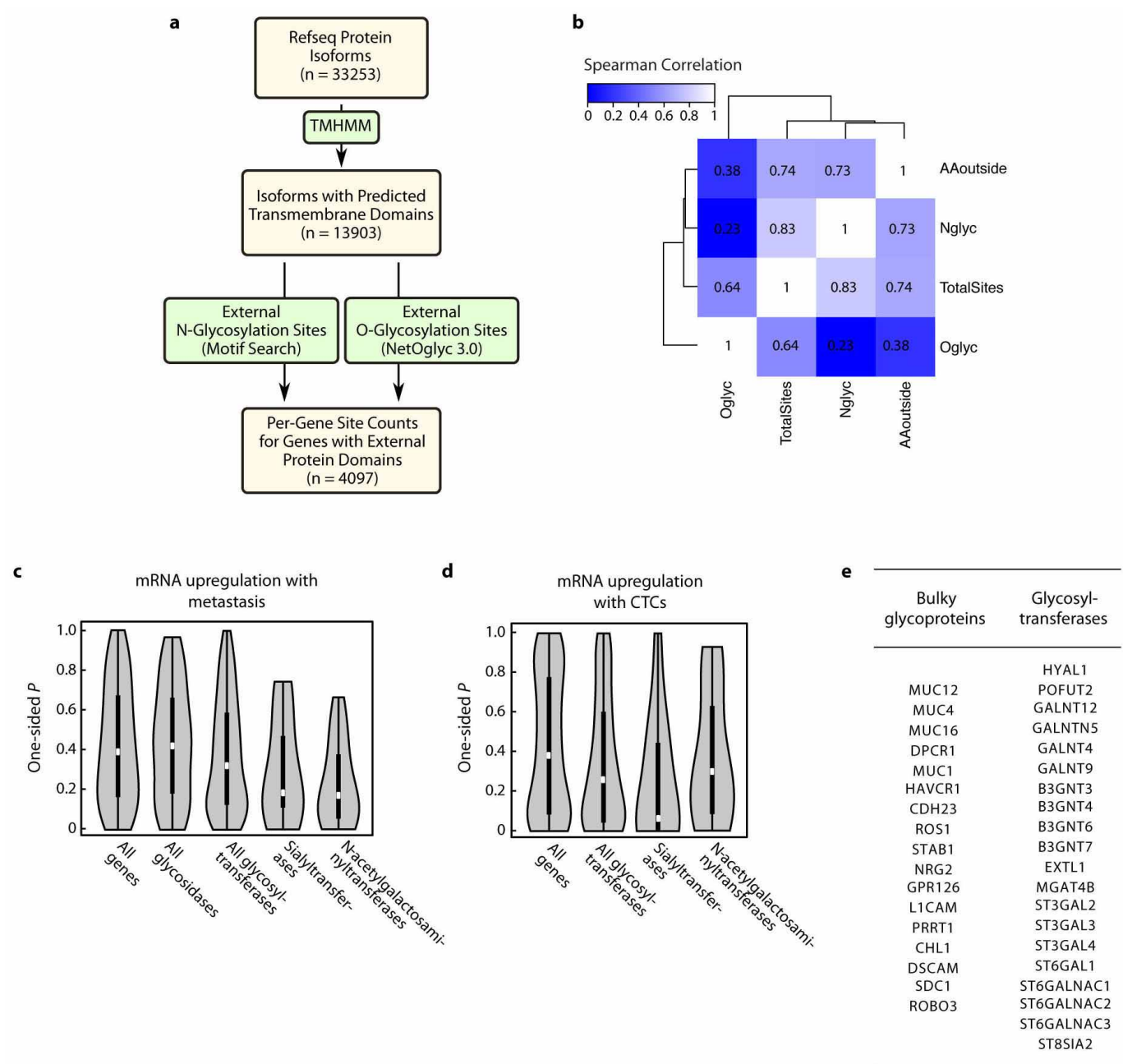
Assessment of fibronectin-fibrillogenesis. Human recombinant fibronectin was labelled with N-hydroxysuccinimide Alexa568 (Invitrogen) according to manufacturer's protocol and dialysed extensively in PBS. Conversion of soluble, fluorescently labelled fibronectin from the growth media into insoluble fibrils was imaged according to published protocol³⁴. Briefly, MCF10A complete growth media

was prepared with donor horse serum that was depleted of fibronectin using gelatin Sepharose 4B (GE Healthcare). MCF10A cells were plated in the depleted media on fibronectin-conjugated glass coverslips and incubated the next day in 10 μ g ml⁻¹ labelled fibronectin for one hour. Cells were quickly rinsed in PBS, fixed in 4% paraformaldehyde, and imaged at random on a spinning disk confocal.

Isolation and gene expression profiling of CTCs. Twenty CTC samples were isolated from the blood of 18 metastatic breast cancer patients as previously described³⁵. Briefly, whole blood was subjected to EpCAM-based immunomagnetic enrichment followed by fluorescence-activated cell sorting of CTCs defined as nucleated, EpCAM-positive, CD45-negative cells. CTCs were sorted directly onto lysis buffer (Taqman PreAmp Cells-to-Ct kit, Life Technologies). cDNA of target genes were pre-amplified (14 cycles) and measured via qPCR analysis. The mean Ct for *ACTB* and *GAPDH* was used for normalization to calculate relative gene expression (Δ Ct). Studies involving CTCs were approved by the UCSF Committee on Human Research. Samples were obtained with IRB approved consent from all patients.

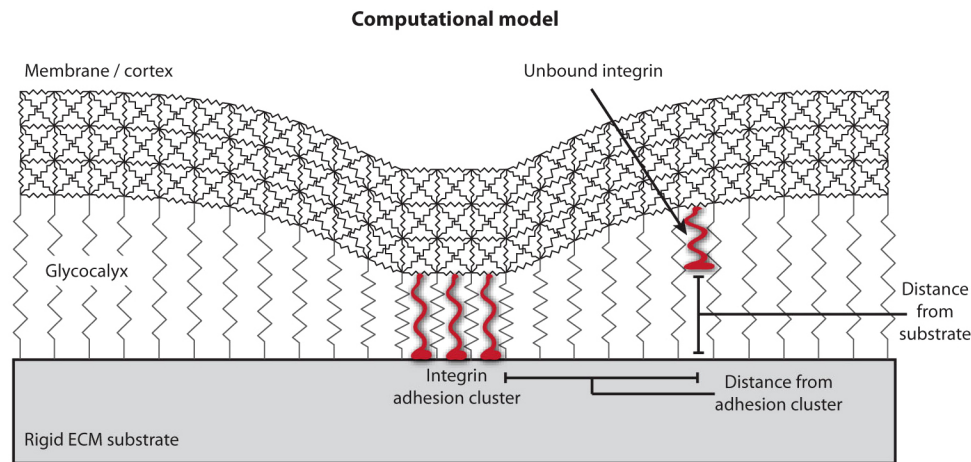
Immunofluorescence labelling of CTCs. CTC samples were isolated from the blood of three metastatic breast cancer patients as described for gene expression profiling. Isolated CTCs were mounted and fixed on poly-L-lysine-coated slides and labelled with FITC-conjugated MUC1 mAb (Clone HPMV). As a control, purified white blood cells from the same patients were prepared similarly, and their immunofluorescence was compared to CTC samples.

Statistics. Statistical significance of experimental data sets was determined by Student's *t*-test after confirming that the data met appropriate assumptions (normality, homogenous variance and independent sampling). Statistical analyses of microarray gene expression data sets are described in detail in Supplementary Note 1. All public microarray data were downloaded from the NCBI Gene Expression Omnibus website and analysed using custom R scripts (all Perl, PHP and R scripts used in this work are available on request).



Extended Data Figure 1 | Large-scale gene expression analysis reveals increased expression of genes encoding bulky glycoproteins and glycan-modifying enzymes in primary tumours of patients with disseminated disease. **a**, Bioinformatics pipeline to estimate the extracellular bulkiness of a protein from its corresponding amino acid sequence. For each isoform sequence, the transmembrane and extramembrane domains were identified using a hidden Markov model (TMHMM). A combination of motif searches and neural network prediction then identified likely N- and O-glycosylation sites within each sequence. Isoform-level bulkiness estimates were generated by summing the number of predicted N- and O-glycosylation sites located within the extramembrane regions of the isoform. **b**, Heat map depicting the pairwise spearman correlation coefficients calculated by comparing all per-gene estimates of the total number of extra-membrane amino acids (AAoutside), N-glycosylation sites (Nglyc), O-glycosylation sites (Oglyc), and the overall bulkiness measure (total sites; for example, the sum of extra-membrane N- and O-glycosylation sites). Correlation coefficients relating the corresponding gene-wise measures are listed in the corresponding cells and depicted on a colour scale, where white corresponds to perfect correlation ($\rho = 1$), and the dendrograms indicate the overall relationship between the parameters, estimated by Euclidean distance. High correlation coefficients indicate that

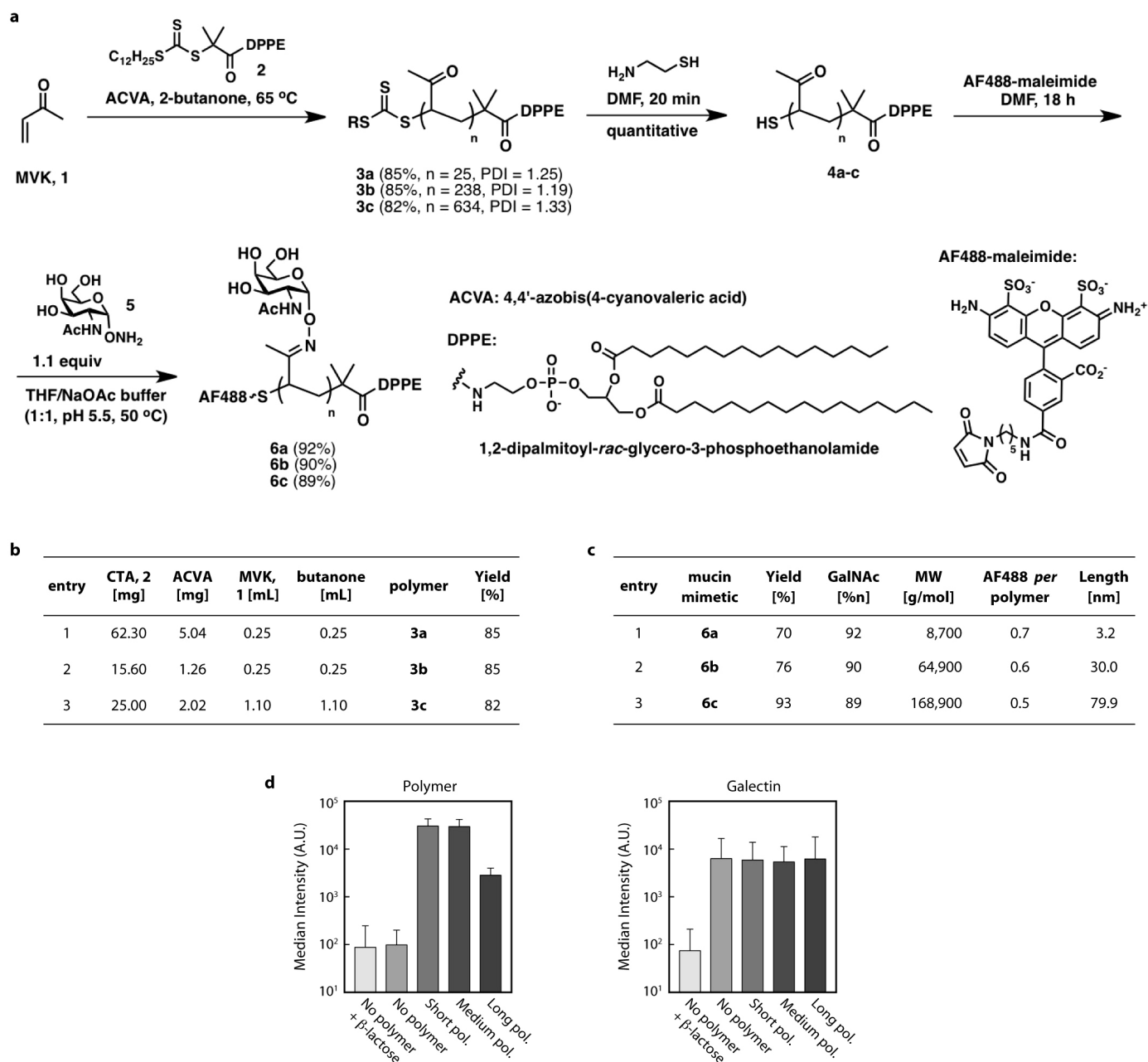
gene-wise estimates of the compared parameters are similarly ranked (for example, genes with high values of X also tend to have high values of Y). The data indicate that the number of extracellular N-glycosylation sites and O-glycosylation sites identified within a gene are only weakly correlated, and neither dominates the total number of sites estimated per gene. **c**, Violin plots contrasting the distributions of gene-wise one-sided P values (y axis) quantifying evidence for transcriptional upregulation of glycosidases and glycosyltransferases, and subsets of glycosyltransferases (sialyltransferases and N-acetylgalactosaminyltransferases) with the full distribution. White dots and thick black lines indicate the median and interquartile range of the gene-wise P -value distribution among category members, and the width of the violin along the y axis indicates the density of the corresponding values. P values are derived from comparisons of expression levels in primary tumours of patients with or without distant metastases using a t -test. Indicated P values were estimated using a one-sided Kolmogorov-Smirnov test. **d**, Violin plots quantifying transcriptional upregulation of glycan-modifying enzymes in primary tumours of patients presenting with circulating tumour cells compared to tumours without detectable circulating tumour cells. **e**, Table of bulky glycoproteins and potential bulk-adding glycosyltransferases whose expression is upregulated in tumours that present with circulating tumour cells.



Extended Data Figure 2 | Computational model of the cell-ECM interface.

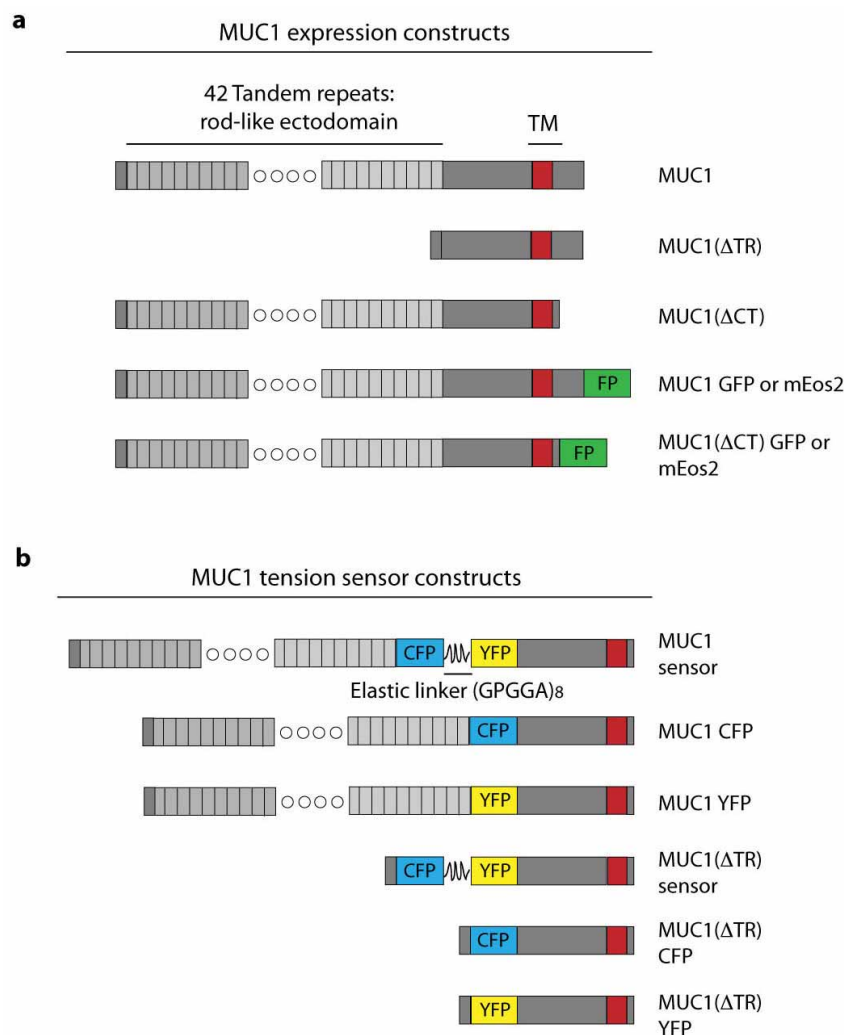
Schematic of an integrated model that describes how the physical properties of the glycocalyx influence integrin-ECM interactions. The cell surface is modelled as a three-dimensional elastic plate; the ECM as a rigid substrate underneath the cell surface; and the glycocalyx as a repulsive potential between the plate and substrate. To compute stress-strain behaviour, the model is discretized using the three-dimensional lattice spring method, the cross-section

of which is depicted above. Integrins are tethered to the cell surface and their distance-dependent binding to the ECM-substrate is calculated according to the Bell model. To calculate integrin-binding rate as a function of lateral distance from an adhesion cluster, an adhesion cluster is first constructed by assembling a 3×3 bond structure. The rates for additional integrin-ECM bonds then are computed at various distances from the cluster.



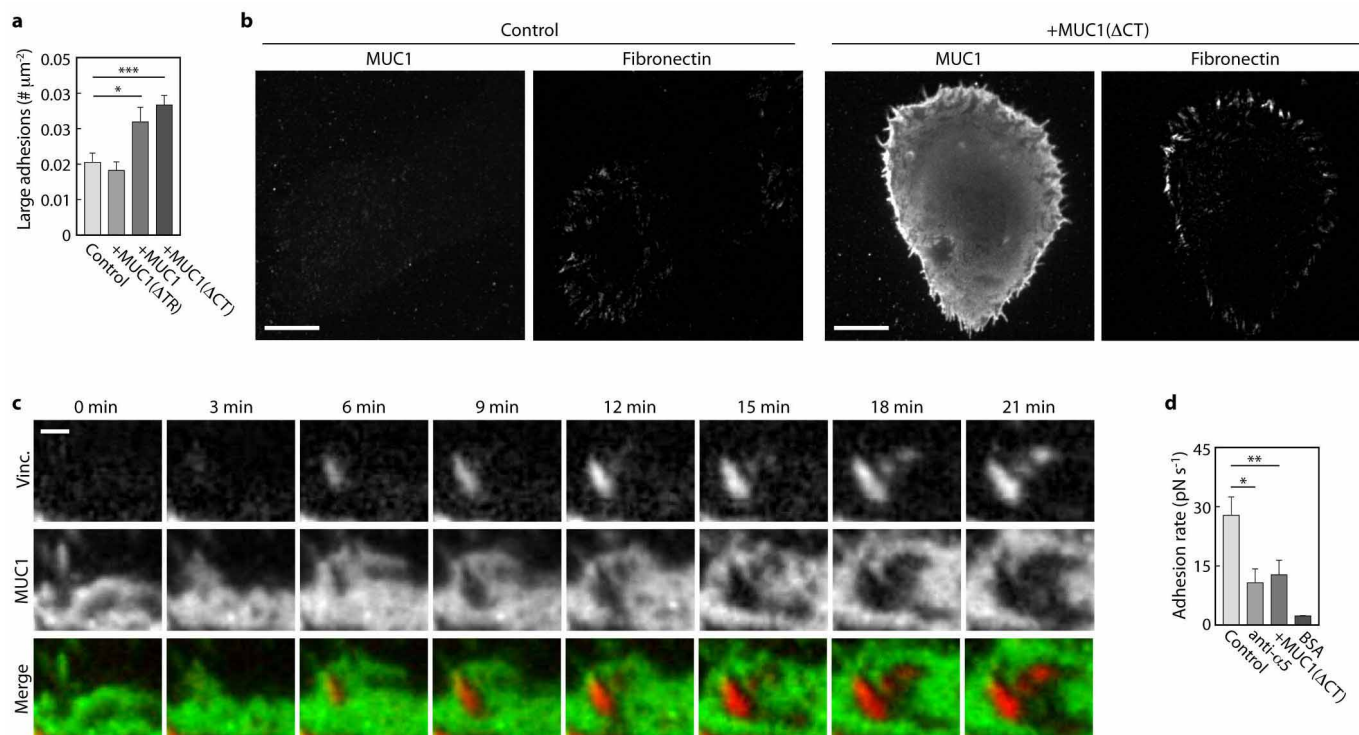
Extended Data Figure 3 | Synthesis and characterization of glycoprotein mimetics. **a**, Scheme for synthesis of lipid-terminated mucin mimetics labelled with Alexa Fluor 488 (AF488). **b**, Reagents and yields for the synthesis of polymers 3a–c. **c**, Characteristics of polymers 6a–c based on ^1H NMR spectra. Glycoprotein mimetics were engineered to have minimal biochemical interactivity with cell surface lectins. **d**, Flow cytometry results quantifying

incorporation of polymer on the surface of mammary epithelial cells (left) and binding with recombinant Alexa568-labelled galectin-3 with or without competitive inhibitor, β -lactose (right). Although a weak affinity between galectin-3 and the pendant N-acetylgalactosimes has previously been reported, the results suggest that incorporation of polymer does not significantly change the affinity of the cell surface for lectins.



Extended Data Figure 4 | MUC1 expression constructs. **a**, Schematic of MUC1 expression constructs. Full-length MUC1 consists of a large ectodomain with 42 mucin-type tandem repeats, a transmembrane domain, and short cytoplasmic tail. The tandem repeats and cytoplasmic tail are deleted in MUC1(ΔTR) and MUC1(ΔCT), respectively. For fluorescent protein fusions, mEmerald (GFP) and mEOS2 are fused to the C terminus of

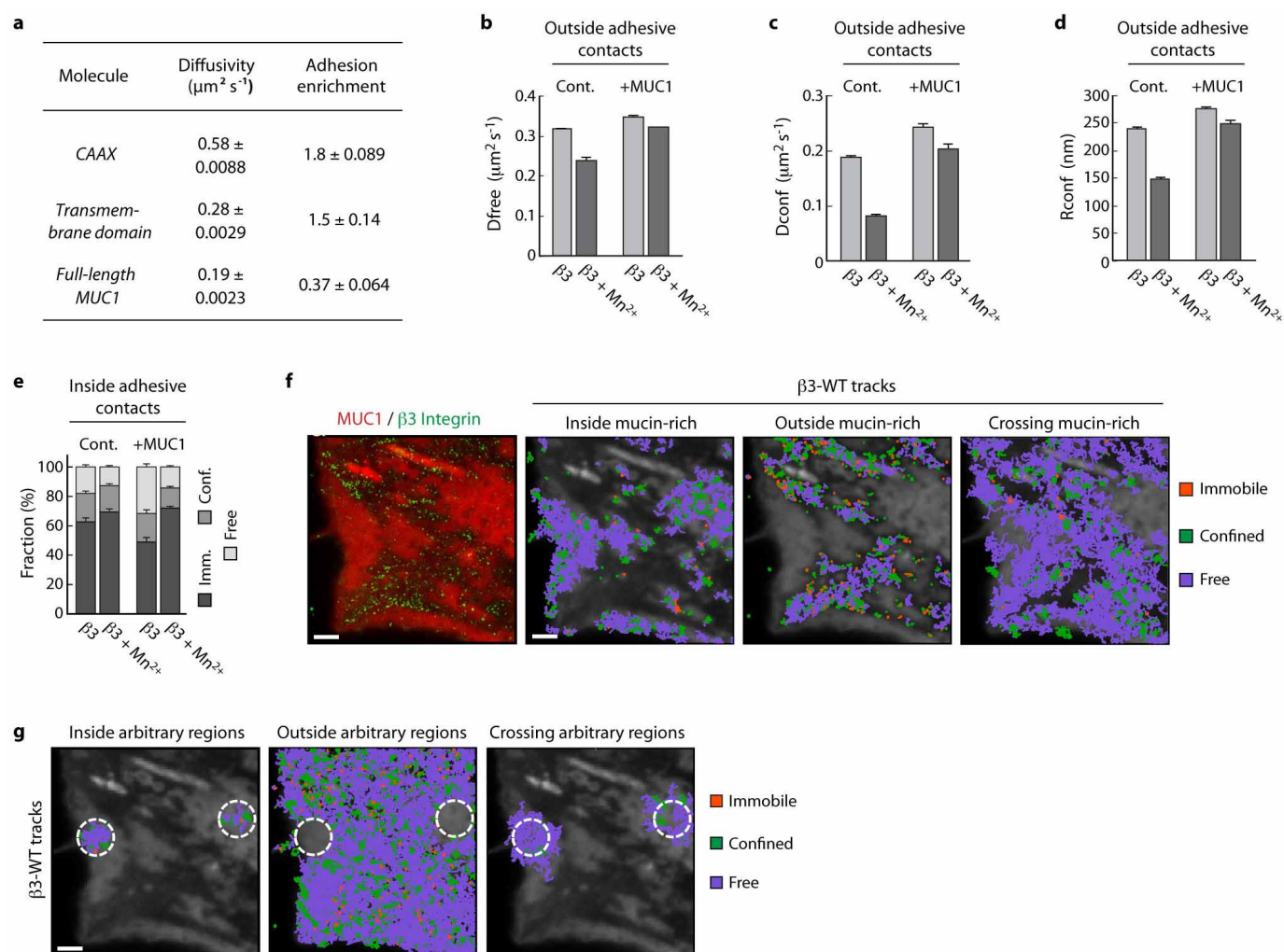
full-length MUC1 or MUC1(ΔCT). **b**, Schematic of MUC1 strain sensor and control constructs. Cysteine-free mTurquoise2 (CFP), Venus (YFP), or a FRET module consisting of the fluorescent proteins separated by an elastic linker (8 repeats of GPGGA) are inserted into the MUC1 ectodomain adjacent to the MUC1 tandem repeats. The mucin tandem repeats are deleted in ectodomain-truncated variants (ΔTR).



Extended Data Figure 5 | MUC1-mediated adhesion formation.

a, Quantification of the average number of large adhesions, greater than $1 \mu\text{m}^2$, per area of cell in control epithelial cells (Control) and those ectopically expressing ectodomain-truncated MUC1 (+MUC1(Δ TR)), wild-type MUC1 (+MUC1), or cytoplasmic-tail-deleted MUC1 (+MUC1(Δ CT)). Results are the mean \pm s.e.m. of three separate experiments. **b**, Fluorescence micrographs showing immuno-labelled MUC1 and fluorescently labelled fibronectin fibrils in control and MUC1-expressing epithelial cells. Soluble, labelled fibronectin in the growth media was deposited by cells at sites of cell-matrix

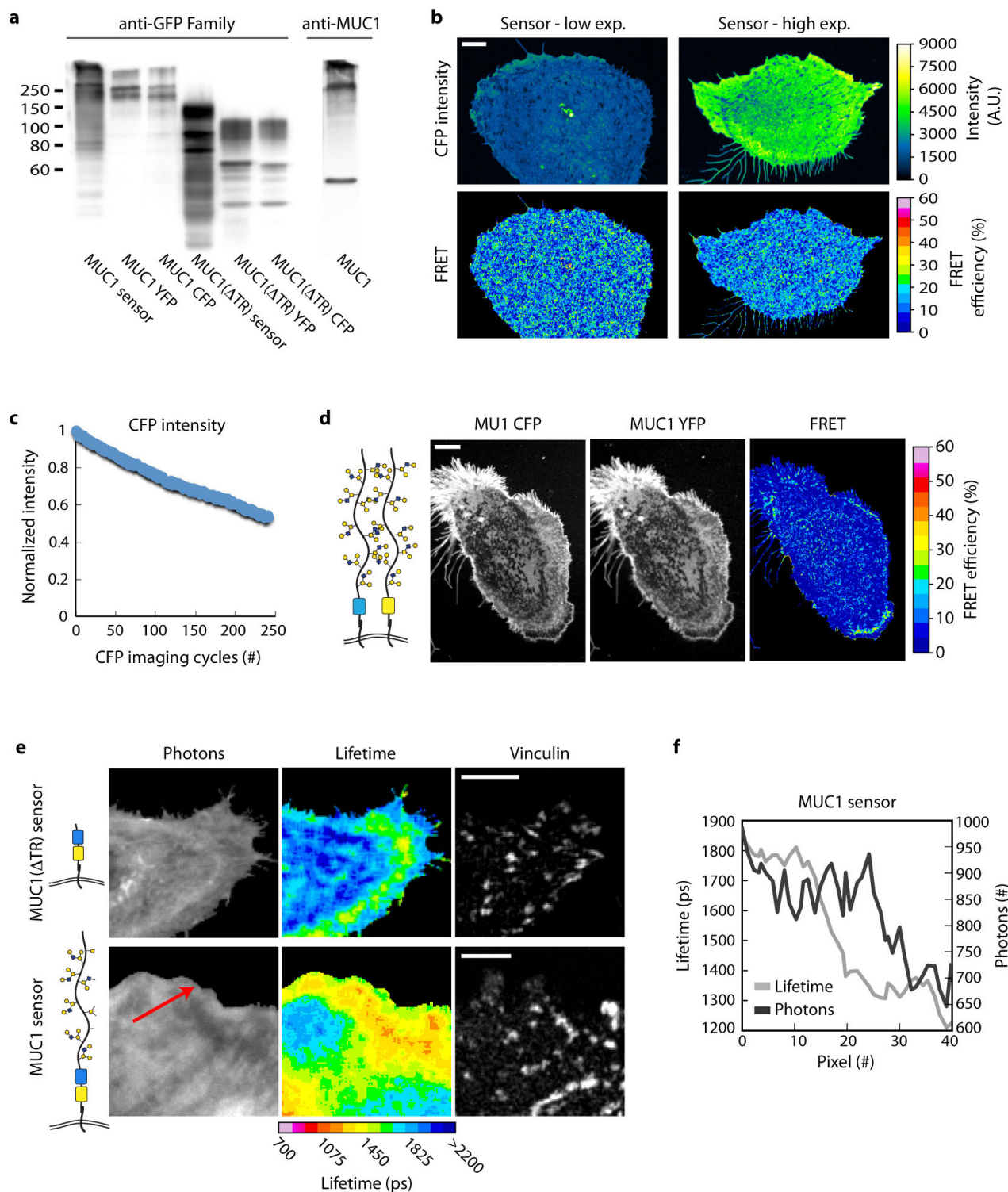
adhesion. Binding of soluble fibronectin to MUC1 was not detected. Scale bar, $10 \mu\text{m}$. **c**, Time lapse images of MUC1-YFP and vinculin-mCherry, showing the dynamics of adhesion assembly (Vinc.) and MUC1 patterning (MUC1). Scale bar, $1 \mu\text{m}$. **d**, Rate of adhesion measured with single cell force spectroscopy of control (Cont.), α_5 integrin-blocked (anti- α_5), and MUC1-expressing cells (+MUC1) to fibronectin-coated surfaces and control cells to BSA-coated surfaces (BSA). Results are the mean \pm s.e.m. of at least 15 cell measurements. Statistical significance is given by * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.



Extended Data Figure 6 | $\beta 3$ integrin mobility in MUC1-expressing cells.

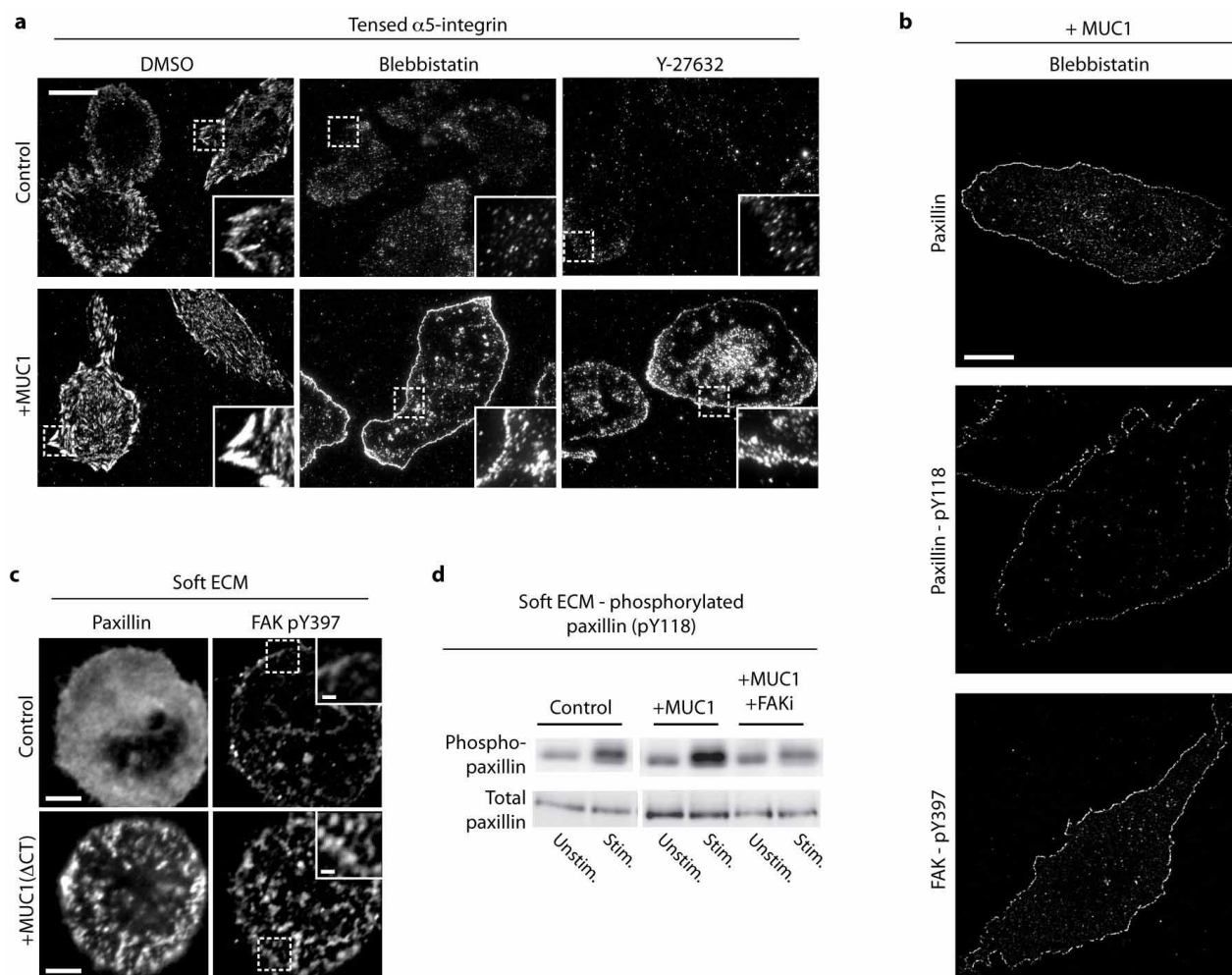
a, Molecular diffusivity and adhesion enrichment measured with sptPALM in mouse embryonic fibroblasts (MEFs). Adhesion enrichment is reported as the ratio of the number of molecules detected inside focal adhesions per unit area to the number of molecules detected outside focal adhesions per unit area. **b**, Mean diffusion coefficients measured for freely diffusive $\beta 3$ integrin tracks outside of adhesive contacts in control (Cont.) and MUC1-transfected (+MUC1) MEFs with and without Mn^{2+} to activate $\beta 3$. **c**, Mean diffusion coefficients measured for confined $\beta 3$ integrin tracks outside of adhesive contacts in MEFs with and without Mn^{2+} . **d**, Mean radius of confinement measured for confined $\beta 3$ integrin tracks outside of adhesive contacts in MEFs with and without Mn^{2+} . **e**, Fraction of immobilized (Imm.), confined (Conf.), and freely diffusive (Free) $\beta 3$ integrins inside of adhesive contacts in control and MUC1-transfected MEFs with and without Mn^{2+} treatment. **f**, From left to

right, panels show GFP-tagged wild-type MUC1 (red) and positions of individual $\beta 3$ integrins (green) in MEFs without Mn^{2+} treatment (left panel) and individual integrin trajectories recorded with sptPALM within MUC1-rich regions, outside MUC1-rich regions, and that cross MUC1 boundaries (scale bar, $2 \mu\text{m}$). The ratio of integrins crossing out versus crossing in the MUC1 boundaries per cell is close to one (1.0 ± 0.1 , $n = 9$ cells, 4,145 trajectories) showing that the flux of free diffusing integrins crossing in or out the mucin region is the same. **g**, From left to right, panels show integrin trajectories within an arbitrary region drawn in a MUC1-rich area (dashed white circles), outside of the circled region, and that cross the circled region (scale bar, $2 \mu\text{m}$). The ratio of integrins crossing the MUC1-rich boundaries versus the fictive boundaries per cell is close to one (1.2 ± 0.2 , $n = 9$ cells, 9,321 trajectories), showing that the MUC1-adhesive zone boundary does not affect the diffusive crossing of integrins. For all bar graphs, results are the mean \pm s.e.m.



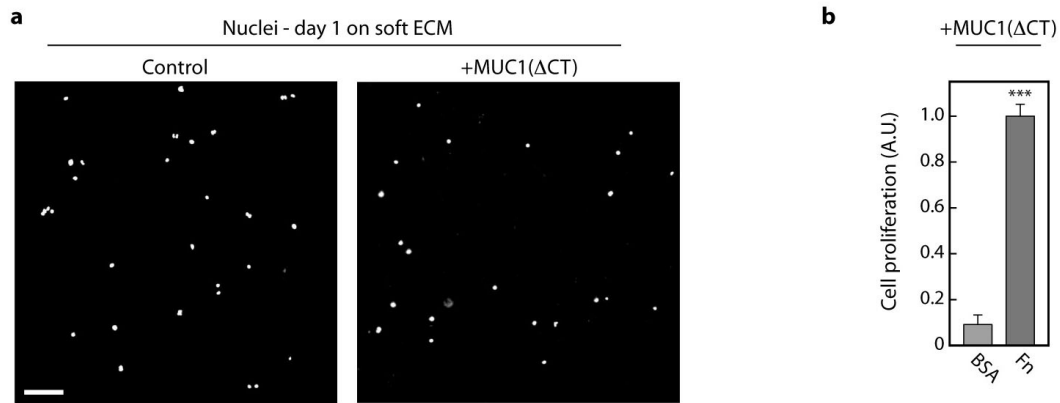
Extended Data Figure 7 | MUC1 strain gauge. **a**, Western blot of indicated construct expressed in HEK 293T cells and probed with anti-GFP family antibody, or full-length MUC1 construct expressed in HEK 293T cells and probed with an antibody against the MUC1 tandem repeats. **b**, Pseudo-coloured images showing similar FRET efficiencies measured by the photobleaching FRET method for mammary epithelial cells (MECs) expressing low (Low) and high (High) levels of the sensor construct. Scale bar, 5 μ m. **c**, Plot showing the level of CFP bleaching per CFP imaging cycle in MECs. **d**, Control images showing minimal intermolecular FRET in MECs expressing similar levels of both MUC1 CFP and MUC1 YFP. **e**, Micrographs showing the emitted

photons from CFP and their fluorescence lifetimes in MECs expressing ectodomain-truncated (MUC1(Δ TR) sensor) or full-length MUC1 strain sensors (MUC1 sensor). Shorter lifetimes are indicative of higher energy transfer between the CFP donor and YFP acceptor, and thus closer spatial proximity of the donor and acceptor (scale bar, 10 μ m). **f**, Representative profile of CFP lifetimes and emitted photons of the full-length MUC1 sensor along the red line in panel **e**. Pixels 0 and 40 correspond to the base and tip of the arrow, respectively. A drop in fluorescence lifetime (Lifetime) is often observed before the drop in MUC1 molecular density (Photons) as an adhesive zone is approached.



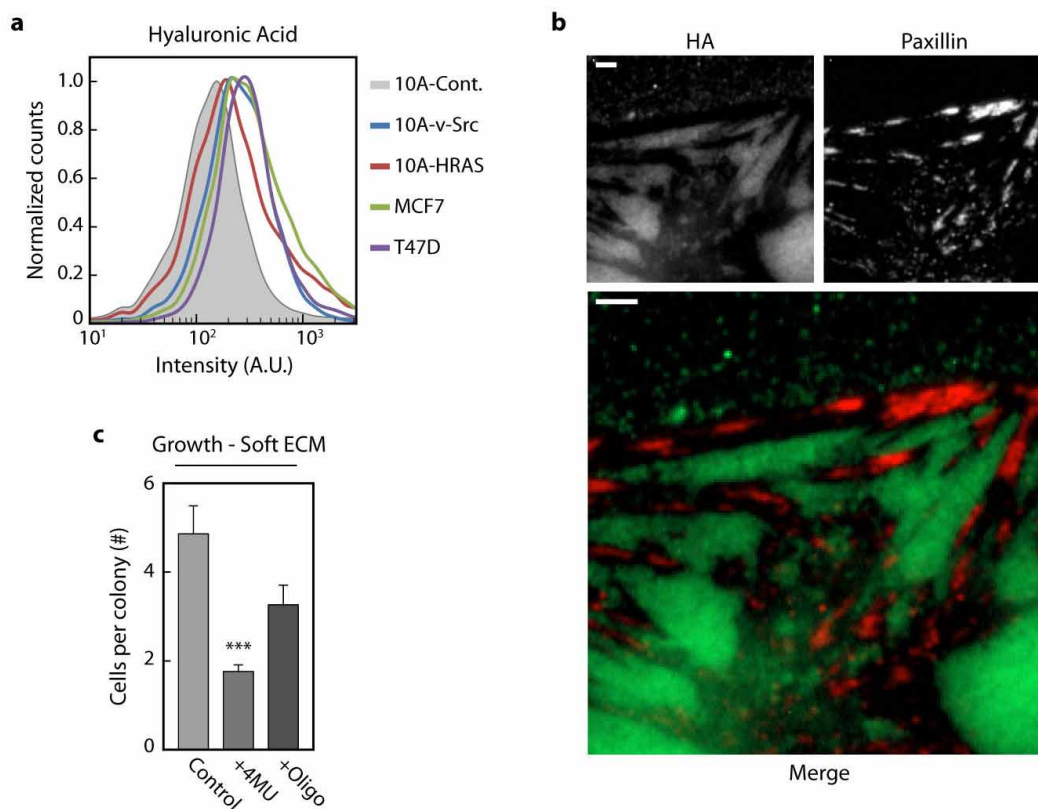
Extended Data Figure 8 | Tension-dependent integrin activation and focal adhesion assembly in MUC1-expressing cells. **a**, Fluorescence micrographs of fibronectin-crosslinked $\alpha 5$ integrin in control and MUC1-expressing mammary epithelial cells (MECs) treated with solvent alone (DMSO), myosin-II inhibitor (blebbistatin; 50 μ M), or Rho kinase inhibitor (Y-27632; 10 μ M) for 1 h and detergent-extracted following crosslinking. Only fibronectin-bound integrins under mechanical tension are crosslinked and visualized following detergent extraction (scale bar, 15 μ m). **b**, Fluorescence micrographs showing formation of myosin-independent adhesion complexes in MUC1-expressing MECs. Cells were pre-treated for 1 h and plated for 2 h in 50 μ M blebbistatin

(scale bar, 10 μ m). **c**, Fluorescence micrographs of paxillin-mCherry and immuno-labelled activated FAK (pY397) in control and MUC1(Δ CT) expressing MECs plated on compliant fibronectin-conjugated hydrogels ($E = 140$ Pa; scale bar, 3 μ m; ROI scale bar, 0.5 μ m). **d**, Western blots showing phosphorylation of paxillin (pY118) in control and MUC1-expressing MECs on compliant substrates ($E = 140$ Pa) following overnight serum starvation and stimulation with EGF. MUC1-expressing cells treated with a pharmacological inhibitor of focal adhesion kinase (+FAKi) for 1 h before EGF stimulation did not exhibit robust paxillin phosphorylation.



Extended Data Figure 9 | Cell proliferation on soft ECM. **a**, Fluorescence micrographs showing DAPI-stained nuclei of control and MUC1(Δ CT)-expressing MECs after 24 h of plating on soft, fibronectin-conjugated hydrogels ($E = 140$ Pa; scale bar, 250 μ m). The majority of cells plated as single cells, indicating that multi-cell colonies that formed at later time points were largely attributed to cell proliferation. **b**, Quantification of cell proliferation of

MUC1(Δ CT)-expressing epithelial cells on soft hydrogels conjugated with bovine serum albumin (BSA) or fibronectin (Fn). Cells plated similarly on BSA- and Fn-hydrogels, but cell proliferation was significantly enhanced on Fn-hydrogels. Results are the mean \pm s.e.m with statistical significance given by $*P < 0.05$; $**P < 0.01$; $***P < 0.001$.



Extended Data Figure 10 | Hyaluronic acid production by tumour cells promotes cellular growth. **a**, Quantification of hyaluronic acid (HA) cell surface levels on control (10A-Cont.), transformed (10A-v-Src, 10A-HRAS) and malignant (MCF7, T47D) mammary epithelial cells (MECs).

b, Fluorescence micrographs of HA and immuno-labelled paxillin on the v-Src transformed MECs (scale bars, 3 μ m). **c**, Quantification of the number of

v-Src-transformed MECs per colony 48 h after plating on soft polyacrylamide gels (fibronectin-conjugated) and treated with vehicle (DMSO), hyaluronic acid synthesis inhibitor 4-methylumbelliferone (+4MU; 0.3 μ M), or competitive inhibitor HA oligonucleotides (+Oligo; 12-mer average oligonucleotide size; 100 mg ml⁻¹). Results are the mean \pm s.e.m with statistical significance is given by * P < 0.05; ** P < 0.01; *** P < 0.001.

Rapid formation of large dust grains in the luminous supernova 2010jl

Christa Gall^{1,2,3}, Jens Hjorth², Darach Watson², Eli Dwek³, Justyn R. Maund^{2,4}, Ori Fox⁵, Giorgos Leloudas^{2,6}, Daniele Malesani² & Avril C. Day-Jones⁷

The origin of dust in galaxies is still a mystery^{1–4}. The majority of the refractory elements are produced in supernova explosions, but it is unclear how and where dust grains condense and grow, and how they avoid destruction in the harsh environments of star-forming galaxies. The recent detection of 0.1 to 0.5 solar masses of dust in nearby supernova remnants^{5–7} suggests *in situ* dust formation, while other observations reveal very little dust in supernovae in the first few years after explosion^{1,8–10}. Observations of the spectral evolution of the bright SN 2010jl have been interpreted as pre-existing dust¹¹, dust formation^{12,13} or no dust at all¹⁴. Here we report the rapid (40 to 240 days) formation of dust in its dense circumstellar medium. The wavelength-dependent extinction of this dust reveals the presence of very large (exceeding one micrometre) grains, which resist destruction¹⁵. At later times (500 to 900 days), the near-infrared thermal emission shows an accelerated growth in dust mass, marking the transition of the dust source from the circumstellar medium to the ejecta. This

provides the link between the early and late dust mass evolution in supernovae with dense circumstellar media.

We observed the bright ($V \approx 14$) and luminous ($M_V \approx -20$) type II_{in} supernova 2010jl (ref. 16) with the VLT/X-shooter spectrograph covering the wide wavelength range 0.3–2.5 μm . Peak brightness occurred on 2010 October 18.6 UT, and observations were made at nine early epochs and at one late epoch, 26–239 days and 868 days past peak, respectively (Methods, Extended Data Table 1, Extended Data Figs 1–5). Figure 1 shows the intermediate-width components of the hydrogen emission lines of $H\gamma$ $\lambda 4,340.472$ (that is, $H\gamma$ at a wavelength of $\lambda = 4,340.472$ Å) and $P\beta$ $\lambda 12,818.072$ and of the oxygen ejecta emission lines [O I] $\lambda 6,300.304$, $\lambda 6,363.776$ (rest frame). The emission profiles change with time, exhibiting a substantial depression of the red wings and a corresponding blueshift of the centroids of the lines (Extended Data Fig. 6) due to preferential extinction of the emission from the receding material on the far side of the supernova^{13,17,18}. The effect is less pronounced at

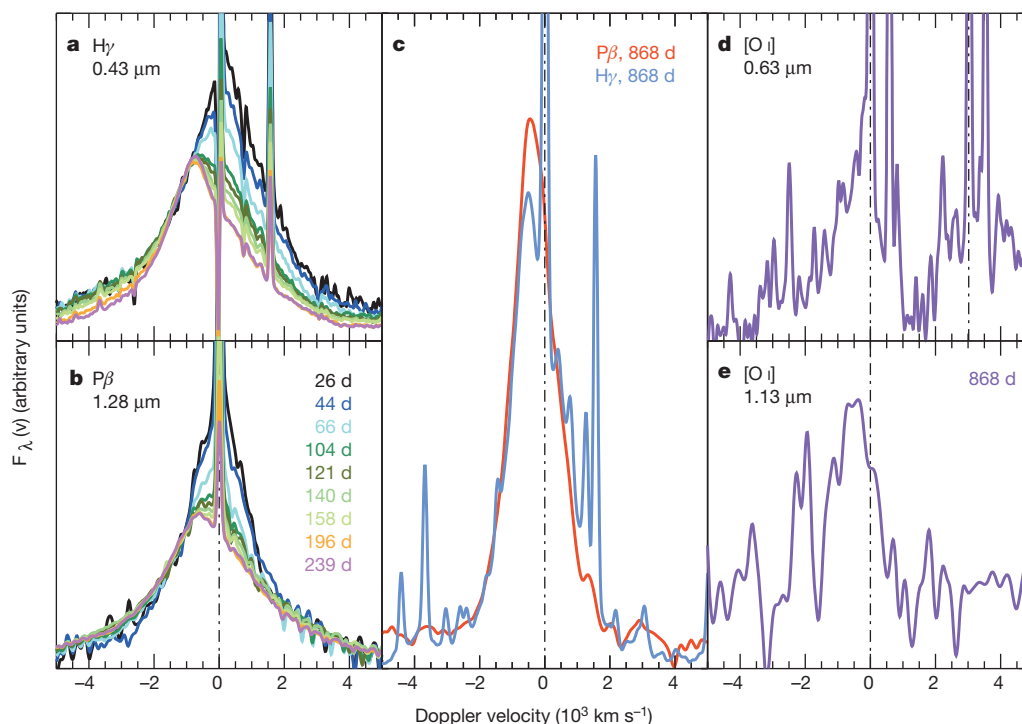


Figure 1 | Evolution of the hydrogen and oxygen line profiles in the spectrum of SN 2010jl. Line profiles for $H\gamma$ $\lambda 4,340.472$ (a) and $P\beta$ $\lambda 12,818.072$ (b) for epochs from 26 days to 239 days and $H\gamma$ and $P\beta$ at 868 days (c). d, The [O I] $\lambda 6,300.304$, $\lambda 6,363.776$ doublet (zero velocity set at $\lambda 6,300.304$). e, The

[O I] $\lambda 11,297.68$ line. The dashed-dotted lines in all panels denote zero velocity, at redshift $z = 0.01058$, as determined from narrow emission lines in the spectrum.

¹Department of Physics and Astronomy, Aarhus University, Ny Munkegade 120, DK-8000 Aarhus C, Denmark. ²Dark Cosmology Centre, Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, DK-2100 Copenhagen Ø, Denmark. ³Observational Cosmology Lab, Code 665, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. ⁴Astrophysics Research Centre School of Mathematics and Physics, Queen's University Belfast, Belfast BT7 1NN, UK. ⁵Department of Astronomy, University of California, Berkeley, California 94720-3411, USA. ⁶The Oskar Klein Centre, Department of Physics, Stockholm University, Albanova University Centre, 10691 Stockholm, Sweden. ⁷Departamento de Astronomía, Universidad de Chile, Camino del Observatorio 1515, Santiago, Chile.

longer wavelengths, as expected if the attenuation of the lines is due to dust extinction, and rules out that the blueshifts are due to electron scattering¹⁴ (Supplementary Information). The early epoch hydrogen lines have a Lorentzian half-width at half-maximum (HWHM) in the range $1,000\text{--}2,000\text{ km s}^{-1}$. The middle and right panels of Fig. 1 show that the line profiles at the late epoch are narrower ($\text{HWHM} \approx 800 \pm 100\text{ km s}^{-1}$) and also exhibit blueshifts of the oxygen lines, which indicates that ejecta material is involved in the dust formation at this stage.

Figure 2 shows the temporal evolution of the inferred extinction A_λ , as derived from the attenuation of emission lines in the early spectra. We calculated the extinction from the ratios of the integrated line profiles at each epoch. We assume that the first epoch at 26 days past peak is nearly unextinguished and use it as a reference. The monotonic increase of the extinction as a function of time indicates continuous formation of dust. The extinction at 239 days is $A_V \approx 0.6\text{ mag}$. Interestingly, the shape of the normalized extinction curve shows no substantial variation with time. Scaling and combining the data from the

eight individual early epochs allowed us to produce the first directly measured, robust extinction curve for a supernova. The extinction curve is shallow, with $R_V = A_V/E(B - V) \approx 6.4$, and can be represented by a mix of grey extinction dust grains (A_λ is a constant) and either standard Small Magellanic Cloud or Milky Way extinction¹⁹. The extinction contribution of the grey dust is 40% in the V band. We fitted several dust models to the extinction curve using amorphous carbon dust characterized by a power-law grain size distribution²⁰ with slope α , and minimum and maximum grain radii ($a_{\min} < a_{\max}$) in the interval $[0.001, 5.0]\text{ }\mu\text{m}$.

Figure 3 shows the resulting confidence interval for the two parameters a_{\max} and α around the best-fit values of $a_{\min} = 0.001\text{ }\mu\text{m}$, $a_{\max} = 4.2\text{ }\mu\text{m}$ and $\alpha = 3.6$. It is evident that only size distributions extending to grain radii that are significantly larger than that of Milky Way interstellar-medium^{21,22} dust ($\geq 0.25\text{ }\mu\text{m}$) can reproduce the supernova extinction curve (Fig. 2). The 2σ lower limit on the maximum grain size is $a_{\max} > 0.7\text{ }\mu\text{m}$. We cannot perform a similar analysis of the late epoch because the intrinsic line profile at this epoch is unknown and is likely to be strongly affected by extinction¹². However, we note that the blueshift velocities change little with wavelength (Extended Data Fig. 6), suggestive of large grains also at this epoch.

Figure 4 illustrates the continuous build-up of dust as a function of time. The increasing attenuation of the lines is accompanied by increasing emission in the near-infrared (NIR) spectra, from a slight excess over a supernova blackbody fit at early times to total dominance at the late epoch. We fitted the spectra with black bodies, which for the NIR excess yield a constant blackbody radius of $(1.0 \pm 0.2) \times 10^{16}\text{ cm}$ at the early epochs, and a temperature that declines from $\sim 2,300\text{ K}$ to $\sim 1,600\text{ K}$ from day 26 onwards. At the late epoch, we obtain a blackbody radius of $(5.7 \pm 0.2) \times 10^{16}\text{ cm}$ and a temperature of $\sim 1,100\text{ K}$. The high temperatures detected at the early epochs suggest that the NIR excess is due to thermal emission from carbonaceous dust, rather than silicate dust, which has a lower condensation temperature of about $1,500\text{ K}$ (ref. 1). The high temperatures rule out suggestions that the NIR emission is due to pre-existing dust or a dust echo¹¹ (Extended Data Figs 7 and 8, and Supplementary Information). Fitting the NIR excess with a modified black body, assuming the grain composition found in our analysis of the extinction curve (Fig. 3), gives a dust temperature similar to the black-body temperature, which is at all epochs (and at all dust compositions considered) larger than $1,000\text{ K}$. The dust masses inferred from the extinction and NIR emission agree very well. The inferred amount of dust at the late epoch (868 days) is $\sim 2.5 \times 10^{-3} M_\odot$ (where M_\odot is the mass of the Sun) if composed of carbon, but could be up to an order of magnitude larger for silicates (Methods). Our results indicate accelerated dust formation after several hundred days. SN 2010j will

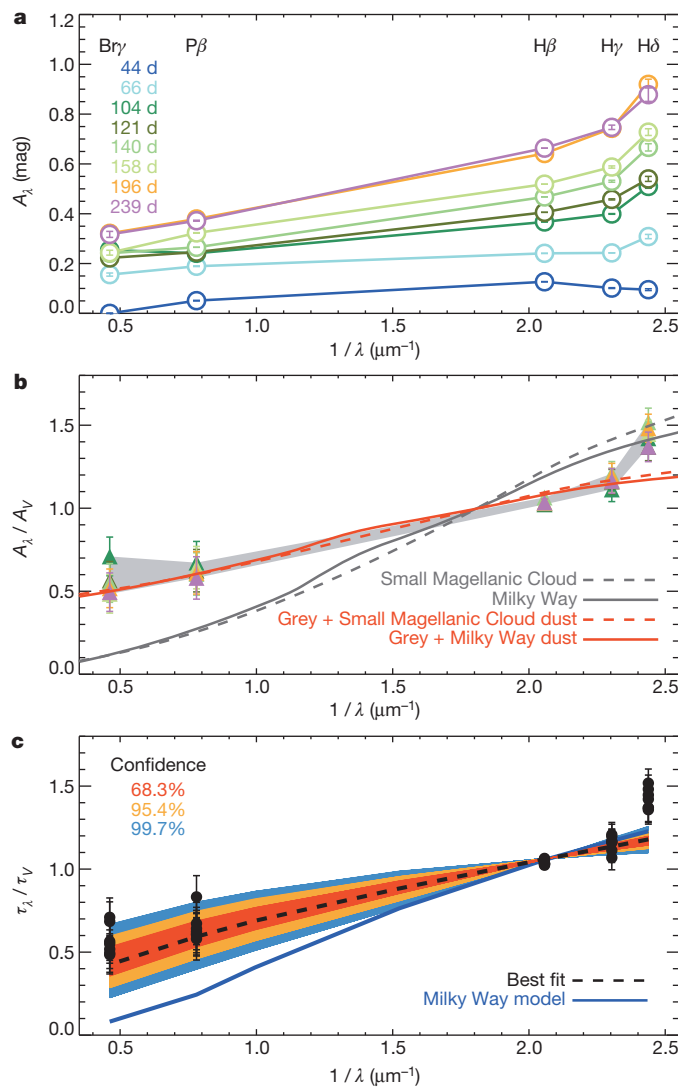


Figure 2 | Supernova dust extinction curves. **a**, The evolution of the extinction A_λ of the hydrogen lines (open circles with standard deviations; see Methods). The solid lines represent the (linearly interpolated) extinction curves. **b**, The grey-shaded area represents the range of extinction curves relative to A_V (filled triangles with error bars). Grey curves are the Small Magellanic Cloud and Milky Way extinction curves, while the red curves include a grey component (Methods). **c**, Fits to the optical depth within the 1σ , 2σ and 3σ (68.3%, 95.4% and 99.7%) confidence intervals (Methods). Dashed and solid curves are models with 'best fitting' and Milky Way parameters, respectively.

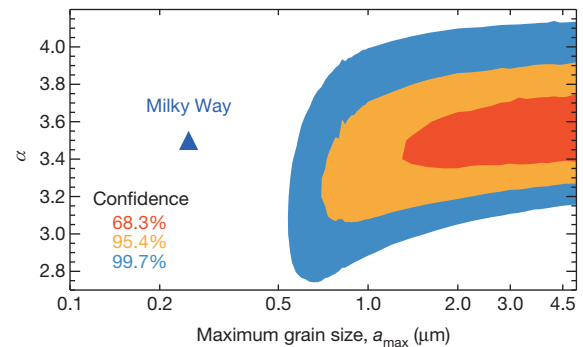


Figure 3 | Maximum grain size and slope of the grain size distribution. Confidence contours, as constrained by the normalized optical depth $\tau(\lambda)$ (see Fig. 2). The most favourable power-law models lie within a parameter range for α (the power-law slope of the grain size distribution) between about 3.4 and 3.7 and require large grains of $a_{\max} \geq 1.3\text{ }\mu\text{m}$ (1σ). The confidence limits are as in Fig. 2. Even at the 3σ confidence limit the maximum grain size is larger ($a_{\max} \geq 0.5\text{ }\mu\text{m}$) than Milky Way maximum grain sizes for a power-law model ($a_{\max} \approx 0.25\text{ }\mu\text{m}$) (ref. 20), or more sophisticated models^{21,22}.

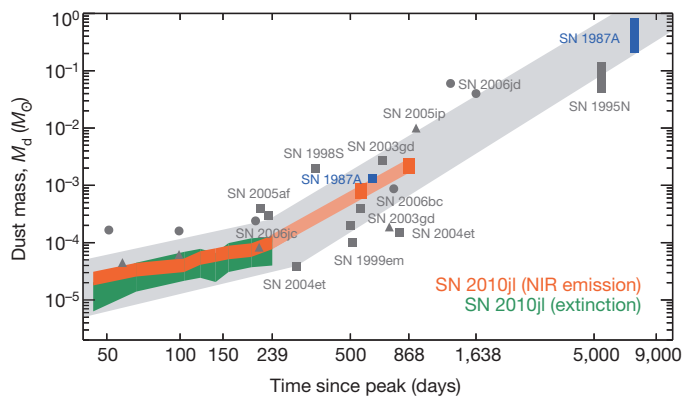


Figure 4 | Temporal evolution of the dust mass. Carbon dust masses and standard deviation derived from the extinction (green band) and the NIR emission (red bars and band; see Methods) including a literature data point at 553 days (ref. 12). The light-grey shaded area illustrates the evolution of the early ($M_d \propto t^{0.8}$ at $t < 250$ days) and late ($M_d \propto t^{2.4}$ at $t > 250$ days) stages of dust formation when SN 2010jl switches from circumstellar to ejecta dust formation. The grey and blue symbols correspond to literature data for SN 2005ip (triangles), SN 2006jd (dots), and other supernovae (bars)^{1,5,6,9,10,27,30}. The length of the symbols for SN 1995N and SN 1987A correspond to the quoted dust mass range. For other supernovae the standard deviation is either smaller than the size of the symbols or has not been reported.

contain a dust mass of $\sim 0.5 M_\odot$ similar to that observed in SN 1987A (refs 5, 6), by approximately day 8,000, if the dust production continues to follow the trend depicted in Fig. 4.

The most obvious location for early dust formation is in a cool, dense shell behind the supernova shock^{18,23}, which sweeps up material as it propagates through the dense circumstellar shell surrounding SN 2010jl²⁴ (Supplementary Information). Dust formation in the ejecta is impossible at this stage because the temperature is too high. The postshock gas cools and gets compressed to the low temperatures and high densities necessary for dust formation and gives rise to the observed intermediate width emission lines. By the time of our first observation at 26 days past peak, the supernova blast wave encounters the dense circumstellar shell at a radius of $\sim 2.0 \times 10^{16}$ cm for a blast wave velocity of $\sim 3.5 \times 10^4$ km s⁻¹. As indicated by the blueshifts of the ejecta metal lines (Fig. 1), the accelerated dust formation occurring at later times (Fig. 4) and at larger radius is possibly facilitated by the bulk ejecta material, which travels on average at a velocity of $\sim 7,500$ km s⁻¹ at early epochs (Extended Data Fig. 4).

Our detection of large grains soon after the supernova explosion suggests a remarkably rapid and efficient mechanism for dust nucleation and growth. The underlying physics is poorly understood but may involve a two-stage process governed by early dust formation in a cool, dense shell, followed by accelerated dust formation involving ejecta material. For type IIP supernovae, the growth of dust grains can be sustained up to 5 years past explosion²⁵. The dense circumstellar material around type II supernovae may provide conditions to facilitate dust growth beyond that. The process appears to be generic, in that other type II supernovae, such as SN 1995N, SN 1998S, SN 2005ip and SN 2006jd, exhibited similar observed NIR properties^{8,10,26,27} and growing dust masses, consistent with the trend revealed here for SN 2010jl (Fig. 4). Moreover, it establishes a link between the early small dust masses inferred in supernovae^{1,8,10} and the large dust masses found in a few supernova remnants^{1,5,7}. Large grains ($0.1 \mu\text{m} \leq a_{\text{max}} \leq 4.0 \mu\text{m}$) provide an effective way to counter destructive processes in the interstellar medium²⁸. Indeed, large grains from the interstellar medium have been detected in the Solar System²⁹. Simulations indicate that grains larger than about $0.1 \mu\text{m}$ will survive reverse shock interactions with only a low fraction being sputtered to smaller radii¹⁵. For a grain size distribution of $a_{\text{min}} = 0.001 \mu\text{m}$, $a_{\text{max}} = 4.2 \mu\text{m}$ and $\alpha = 3.6$ (Figs 2 and 3), the mass fraction of grains above $0.1 \mu\text{m}$ is about 80%, that is, the majority of the produced dust mass can be retained.

METHODS SUMMARY

We obtained optical and near-infrared medium-resolution spectroscopy with the European Southern Observatory's Very Large Telescope (VLT)/X-shooter instrument of the bright type II supernova 2010jl at ten epochs between 2010 November 13.4 UT and 2013 March 4.0 UT. The continuum emission of the spectra was fitted with a combination of black-body, modified black-body and host galaxy models, allowing us to quantify the temporal progression of the temperature and radius of the photosphere as well as the temperature and characteristics of the dust forming, which causes conspicuous excess near-infrared emission. We analysed the profiles of the most prominent hydrogen, helium and oxygen emission lines. From Lorentzian profile fits, which are good representations of the emission lines, we measured the blueshifts of the peaks and the HWHM of the lines, and derived the wavelength-dependent attenuation properties of the dust forming at each epoch. The uncertainties were obtained using Monte Carlo calculations by varying the Lorentzian profile parameters. We generated synthetic UVRIJHK light curves and calculated the energy output of the supernova. This, together with calculated dust vaporization radii, temperatures of the dust grains at different distances from the supernova, and the radius evolution of the forward shock, were used to constrain the location of the dust as it formed. Different dust models, characterized by either single grain sizes or a power-law grain-size distribution function and either amorphous carbon or silicates, were fitted to the extinction curves and the near-infrared excess emission. From these fits, we derived the temporal progression of the dust mass of the dust as it formed at each observed epoch.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 March 2013; accepted 29 May 2014.

Published online 9 July 2014.

- Gall, C., Hjorth, J. & Andersen, A. C. Production of dust by massive stars at high redshift. *Astron. Astrophys. Rev.* **19**, 43 (2011).
- Matsuura, M. *et al.* The global gas and dust budget of the Large Magellanic Cloud: AGB stars and supernovae, and the impact on the ISM evolution. *Mon. Not. R. Astron. Soc.* **396**, 918–934 (2009).
- Draine, B. T. in *Cosmic Dust—Near and Far* (eds Henning, T., Grün, E. & Steinacker, J.) Vol. 414, 453–472 (Astronomical Society of the Pacific Conference Series, 2009).
- Dunne, L. *et al.* Herschel-ATLAS: rapid evolution of dust in galaxies over the last 5 billion years. *Mon. Not. R. Astron. Soc.* **417**, 1510–1533 (2011).
- Matsuura, M. *et al.* Herschel detects a massive dust reservoir in supernova 1987A. *Science* **333**, 1258–1261 (2011).
- Indebetouw, R. *et al.* Dust production and particle acceleration in supernova 1987A revealed with ALMA. *Astrophys. J.* **782**, L2 (2014).
- Gomez, H. L. *et al.* A cool dust factory in the Crab nebula: a Herschel study of the filaments. *Astrophys. J.* **760**, 96 (2012).
- Pozzo, M. *et al.* On the source of the late-time infrared luminosity of SN 1998S and other type II supernovae. *Mon. Not. R. Astron. Soc.* **352**, 457–477 (2004).
- Otsuka, M. *et al.* Late-time light curves of type II supernovae: physical properties of supernovae and their environment. *Astrophys. J.* **744**, 26 (2012).
- Stritzinger, M. *et al.* Multi-wavelength observations of the enduring type II supernovae 2005ip and 2006jd. *Astrophys. J.* **756**, 173 (2012).
- Andrews, J. E. *et al.* Evidence for pre-existing dust in the bright type II supernova 2010jl. *Astron. J.* **142**, 45 (2011).
- Maeda, K. *et al.* Properties of newly formed dust grains in the luminous type II supernova 2010jl. *Astrophys. J.* **776**, 5 (2013).
- Smith, N. *et al.* Systematic blueshift of line profiles in the type II supernova 2010jl: evidence for post-shock dust formation? *Astron. J.* **143**, 17 (2012).
- Zhang, T. *et al.* Type II supernova SN 2010jl: optical observations for over 500 days after explosion. *Astron. J.* **144**, 131 (2012).
- Silvia, D. W., Smith, B. D. & Shull, J. M. Numerical simulations of supernova dust destruction. I. Cloud-crushing and post-processed grain sputtering. *Astrophys. J.* **715**, 1575–1590 (2010).
- Newton, J. & Puckett, T. Possible supernova in UGC 5189A. *Centr. Bur. Electron. Telegr.* **2532**, 1 (2010).
- Lucy, L. B., Danziger, I. J., Gouffes, C. & Bouchet, P. in *IAU Colloq. 120: Structure and Dynamics of the Interstellar Medium* (eds Tenorio-Tagle, G., Moles, M. & Melnick, J.) *Lecture Notes in Physics* Vol. 350, 164–179 (Springer, 1989).
- Smith, N., Foley, R. J. & Filippenko, A. V. Dust formation and He II $\lambda 4686$ emission in the dense shell of the peculiar type Ib supernova 2006jc. *Astrophys. J.* **680**, 568–579 (2008).
- Gordon, K. D., Clayton, G. C., Misselt, K. A., Landolt, A. U. & Wolff, M. J. A quantitative comparison of the Small Magellanic Cloud, Large Magellanic Cloud, and Milky Way ultraviolet to near-infrared extinction curves. *Astrophys. J.* **594**, 279–293 (2003).
- Mathis, J. S., Rumpl, W. & Nordsieck, K. H. The size distribution of interstellar grains. *Astrophys. J.* **217**, 425–433 (1977).
- Zubko, V., Dwek, E. & Arendt, R. G. Interstellar dust models consistent with extinction, emission, and abundance constraints. *Astrophys. J., (Suppl.)* **152**, 211–249 (2004).
- Brandt, T. D. & Draine, B. T. The spectrum of the diffuse galactic light: the Milky Way in scattered light. *Astrophys. J.* **744**, 129 (2012).

23. Fox, O. *et al.* Near-infrared photometry of the type II_n SN 2005ip: the case for dust condensation. *Astrophys. J.* **691**, 650–660 (2009).
24. Ofek, E. O. *et al.* SN 2010jl: optical to hard X-ray observations reveal an explosion embedded in a ten solar mass cocoon. *Astrophys. J.* **781**, 42 (2014).
25. Sarangi, A. & Cherchneff, I. The chemically controlled synthesis of dust in type II-P supernovae. *Astrophys. J.* **776**, 107 (2013).
26. Mauerhan, J. & Smith, N. Supernova 1998S at 14 years postmortem: continuing circumstellar interaction and dust formation. *Mon. Not. R. Astron. Soc.* **424**, 2659–2666 (2012).
27. Van Dyk, S. D. Late-time dust emission from the type II_n supernova 1995N. *Astron. J.* **145**, 118 (2013).
28. Jones, A. P. & Nuth, J. A. Dust destruction in the ISM: a re-evaluation of dust lifetimes. *Astron. Astrophys.* **530**, A44 (2011).
29. Frisch, P. C. *et al.* Dust in the local interstellar wind. *Astrophys. J.* **525**, 492–516 (1999).
30. Gallagher, J. S. *et al.* Optical and infrared analysis of type II SN 2006bc. *Astrophys. J.* **753**, 109 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank L. Christensen and T. Frederiksen for advice on data reduction with the X-shooter pipeline and M. Stritzinger and R. Arendt for discussions.

This investigation is based on observations made with ESO Telescopes at the La Silla Paranal Observatory under programme ID numbers 084.C-0315(D) and 087.C-0456(A). C.G. was supported from the NASA Postdoctoral Program (NPP) and acknowledges funding provided by the Danish Agency for Science and Technology and Innovation. G.L. is supported by the Swedish Research Council through grant number 623-2011-7117. A.C.D.-J. is supported by the Proyecto Basal PB06 (CATA), and partially supported by the Joint Committee ESO-Government Chile. The Dark Cosmology Centre is funded by the Danish National Research Foundation.

Author Contributions C.G. and J.H. conducted the observational campaign, reduced and analysed the data and wrote the manuscript. D.W. was the Principal Investigator of the observing programmes and assisted in writing the manuscript. E.D. performed calculations of vaporization radii and assisted in writing the manuscript. O.F. and G.L. assisted in data analysis. J.R.M. helped with the interpretation of the spectra and line profiles. D.M. and D.W. assisted with observations. A.C.D.-J. conducted the observation of the epoch 2 spectrum. All authors were engaged in discussions and provided comments on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.G. (cgal@phys.au.dk).

METHODS

We observed the type IIIn SN 2010jl in UGC 5891A at ten epochs between 2010 November 13.4 UT and 2013 March 4.0 UT, following its discovery on 2010 November 3.5 UT (ref. 31). The supernova was first detected from the All Sky Automated Survey North on 2010 October 9.6 UT and peaked on 2010 October 18.6 UT (ref. 32). The time of explosion is unknown, but we assume a time of rise to peak of about 40 days. We adopt a luminosity distance of $D = 45.7$ Mpc to the supernova, based on our measured redshift of $z = 0.01058$.

The observations were obtained with the X-shooter echelle spectrograph^{33,34} mounted at the Cassegrain focus of the Kueyen unit of the Very Large Telescope (VLT) at the European Southern Observatory (ESO) on Cerro Paranal, Chile. The X-shooter instrument allows for simultaneous spectroscopic observations in three different arms, the ultraviolet and blue (UVB), visual (VIS) and near-infrared (NIR) wavebands, covering the continuous wavelength range of 0.3–2.5 μm . The observations were performed at the parallactic angle, with nodding between exposures along the 11" slit (see Extended Data Table 1 for details). The spectra were obtained under the following conditions: clear sky and in some cases thin cirrus, average seeing of $\sim 0.8''$, and range in air mass of ~ 1.2 –2.0. For the majority of the observations we used slit widths of 1.0" (UVB), and 0.9" (VIS and NIR) giving resolving powers of 5,100 (UVB), 8,800 (VIS) and 5,300 (NIR), except for the second epoch on 2010 December 1.4 UT where, owing to mediocre seeing conditions of around 1.7", we used wider slit widths of 1.6" (UVB), and 1.5" (VIS and NIR), leading to reduced resolving powers of 3,300 (UVB), 5,400 (VIS) and 3,500 (NIR). For all epochs, observations of spectrophotometric standards were performed using a slit width of 5.0".

We used versions 1.5.0 and 2.2.0 of the X-shooter pipeline³⁵ in physical mode to reduce the supernova and the standard star spectra to two-dimensional bias-subtracted, flat-field corrected, order-rectified and wavelength-calibrated spectra in counts. To obtain one-dimensional spectra the two-dimensional spectra from the pipeline were optimally extracted³⁶. Furthermore, the spectra were slit-loss corrected, flux calibrated and corrected for heliocentric velocities. Additionally, telluric corrections were applied. All calibration and correction procedures after the basic pipeline reduction were performed using custom IDL programs. The spectra were corrected for a Galactic extinction along the line of sight to the supernova of $E(B - V)$ of 0.027 mag (ref. 37).

Two-temperature black-body fits. The progressive evolution of the supernova spectra is shown in Extended Data Fig. 1. We fitted the continuum emission of the early epochs (26–239 days) with a combination of two black-body functions. The first black body represents the supernova photosphere, for which we infer a temperature $T_{\text{SN}} \approx 7,300$ K and a photospheric radius decreasing from radius $R_{\text{SN}} \approx 3.2 \times 10^{15}$ cm at 26 days to $R_{\text{SN}} \approx 2.4 \times 10^{15}$ cm at 239 days. The second black-body function accounts for the NIR excess noticeable in the spectra, which we attribute to dust emission.

To properly fit the hot dust emission we therefore used a modified black-body function. The fit to the spectra is computed as

$$F_{\nu}(v) = B_{\nu}(v, T_{\text{SN}}) R_{\text{SN}}^2 / D^2 + N_d / D^2 \int_{a_{\min}}^{a_{\max}} f(a) m(a) \kappa_{\text{abs}}(v, a) B_{\nu}(v, T_{\text{hot}}) da \quad (1)$$

where $B_{\nu}(v, T)$ is the Planck function at temperature $T = T_{\text{SN}}$ for the supernova and $T = T_{\text{hot}}$ for the dust, R_{SN} is the radius of the supernova photosphere, D is the luminosity distance to the supernova, N_d is the total number of dust particles, $m_d(a) = (4\pi/3)\rho a^3$ is the mass of a single dust grain of radius a and $\kappa_{\text{abs}}(v)$ is the dust mass absorption coefficient for an assumed dust composition, that is, amorphous carbon³⁸ and silicates³⁹. The mass density is $\rho = 1.8 \text{ g cm}^{-3}$ for amorphous carbon and $\rho = 3.3 \text{ g cm}^{-3}$ for silicates. We used a power-law grain size distribution function $f(a) \propto a^{-\alpha}$, which is normalized to unity in the interval $[a_{\min}, a_{\max}]$ as $\int_{a_{\min}}^{a_{\max}} f(a) da = 1$.

Extended Data Fig. 2 depicts supernova spectra obtained at 44 days, 196 days and at a late epoch of 868 days. The adopted grain size distribution assumes the parameters α , a_{\min} and a_{\max} from the best-fitting amorphous carbon model obtained from the extinction curves (Figs 2 and 3). To fit the spectrum at 868 days we exchanged the supernova black-body with a power law for the host galaxy continuum emission, expressed as $C_{\text{norm}} \times v^{-1.24}$, where C_{norm} is a normalization constant and the power law exponent resulted from the fit. Additionally, for this epoch we explored two dust compositions, that is, amorphous carbon and silicates, and models with single grain sizes a between 0.001 μm and 5.0 μm , as well as grain-size distribution models varying α between 2.0 and 4.5. We found (1) that amorphous carbon single-grain-size models as well as grain-size distribution models prefer large grains (1–5 μm), (2) that the quality of the fits of silicate models is fairly insensitive to the size of the grains and can accommodate small grains, (3) that we are unable to produce models with temperatures less than $\sim 1,000$ K, and (4) that the inferred dust masses for silicate grains

are typically up to an order of magnitude higher than for amorphous carbon. All spectra are well fitted by a supernova temperature $T_{\text{SN}} \approx 7,300$ K and a dust temperature T_{hot} which decreases from approximately 2,300 K to 1,600 K during the first 239 days, down to approximately 1,100 K at 868 days.

Extended Data Fig. 1 also shows Spitzer/IRAC 3.6 μm and 4.5 μm observations⁴¹. We can fit the 3.6 μm data point with the same modified black-body model as used for the other epochs (grey dotted curve in Extended Data Fig. 1).

Analysis of line profiles. The spectra exhibit a richness of emission lines on top of the continuum, featuring in particular hydrogen and helium lines, which are: H δ λ 4,101.734, H γ λ 4,340.472, H β λ 4,861.35, He I λ 5,875.621, H α λ 6,562.79, He I λ 7,065.2578, P δ λ 10,049.8, He I λ 10,830.199, P γ λ 10,938.17, P β λ 12,818.072 and Br γ λ 21,655.268. The lines have a narrow ($\sim 100 \text{ km s}^{-1}$) velocity component on top of an intermediate-width velocity component. For H δ , H γ , H β , H α , He I λ 5,875.621 and He I λ 10,830.199, the narrow lines exhibit a characteristic P Cygni profile (that is, a blue-shifted absorption and redshifted emission component).

Only a subset of the hydrogen lines is suitable for quantitative extinction studies. We required that the lines exhibit a clear single-peaked intermediate-velocity component across all epochs, which can be well represented by a Lorentzian profile, not necessarily centred at the zero velocity (Extended Data Fig. 3a). None of the He I lines are suitable for extinction studies because they show conspicuous bumps in the wings. Moreover, the wings significantly broaden with time (Extended Data Fig. 3b). The He I λ 10,830.199 line is blended with the P γ λ 10,938.17 line, ruling out both lines for our studies. The P δ λ 10,049.8 line is located at the crossover between the X-shooter VIS and NIR arms, giving rise to unreliable flux calibration and background subtraction.

Some lines show the presence of large velocities. Extended Data Fig. 4 shows H β P Cygni profiles featuring velocities up to $\sim 20,000 \text{ km s}^{-1}$ which arise from the fast expanding thin outer layers of the supernova ejecta. The bulk expansion velocity of the supernova ejecta (corresponding to the minimum of the P Cygni profile) is around $7,500 \text{ km s}^{-1}$. Other lines, for example, H α , are characterized by an underlying broad velocity component. As shown in Extended Data Fig. 5a, we can fit the 26-day H α line with a combination of a broad Gaussian with a full-width at half-maximum of $\sim 5,000 \text{ km s}^{-1}$ and an intermediate Lorentzian with HWHM $\approx 860 \text{ km s}^{-1}$ centred at zero velocity. Although the H α line is often used to demonstrate the effect of dust attenuation of the red wing⁴⁰, it is discarded for our study because of the progressive broadening of the wings (Extended Data Fig. 3b), which prevents a straightforward quantitative analysis.

At 868 days the emission lines no longer exhibit a broad velocity component. The intermediate velocity components of the hydrogen emission lines feature velocities up to $\sim 2,000$ – $3,000 \text{ km s}^{-1}$ similar to the oxygen [O I] λ 6,300.304 and [O I] λ 11,297.68 lines (Fig. 1). The lines are not well represented by Lorentzian profiles (Fig. 1, middle panel). Consequently, the late epoch is not considered for our quantitative extinction studies.

From single Lorentzians fits to H δ , H γ , H β , P β and Br γ , we estimated the Lorentzian HWHM of the intermediate-velocity components of these lines (about $1,500 \pm 200 \text{ km s}^{-1}$). Extended Data Fig. 5b shows that the hydrogen lines (for example, H β) exhibit deviations from symmetry, despite being adequately represented by Lorentzian profiles for our purposes (Extended Data Fig. 3a). We also measured the blueshifts of the peaks (Extended Data Fig. 6).

To obtain the hydrogen line profiles (Fig. 1), the spectrum from each epoch was continuum-subtracted and scaled to the first epoch. The scaling was set by the velocity at which the blue side of the line changed from being extinguished to being unextinguished (between $-1,200 \text{ km s}^{-1}$ and $-1,000 \text{ km s}^{-1}$). This ensures that we measure only the extinguished parts of the lines. The blue unextinguished wings from all epochs coincide. At the late epoch (868 days), H γ was scaled to P β at a velocity of -800 km s^{-1} .

Extinction measurements. Attributing the red depressions to dust, we calculated the extinction (Fig. 2a) from the fitted Lorentzian profiles as $A_{\lambda} = -2.5 \log(I(\lambda, t)/I_{\text{ref}}(\lambda))$, where $I(\lambda, t)$ is the line profile integrated over a velocity range extending from the scaling velocity up to $4,000 \text{ km s}^{-1}$ and $I_{\text{ref}}(\lambda)$ is the integrated line profile from the first epoch which was taken as a reference. We obtained the error bars of A_{λ} (standard deviations) using Monte Carlo calculations by varying the fit parameters of the Lorentzian line profiles within their uncertainties. The error bars reflect the signal-to-noise ratio of the lines and the extent to which they are well represented by Lorentzians. From measurements of A_V ($\lambda_V = 5,505 \text{ \AA}$) and $E(B - V)$ in Fig. 2, we directly infer $R_V = A_V/E(B - V) \approx 6.4$. The wavelength-dependent optical depth is fitted with (1) a phenomenological model based on grey dust plus either Small Magellanic Cloud or Milky Way dust, $A_{\text{SMC,MW}}(\lambda)$, as $A_{\lambda} = A_{\text{grey}} + A_{\text{SMC,MW}}(\lambda)$ (Fig. 2b), and (2) a single-dust model, that is, only carbon dust, which for a shell is

$$\tau(\lambda) = \frac{N_d}{4\pi R^2} \int_{a_{\min}}^{a_{\max}} f(a) m_d(a) \kappa_{\text{ext}}(\lambda, a) da \quad (2)$$

where R is the distance of the cool dense shell (CDS) from the supernova and $\kappa_{\text{ext}}(\lambda, a)$ is the mass absorption coefficient, in this case, for amorphous carbon³⁸. We calculate a grid of models varying the slope α between 0.5 and 4.5, and the lower and upper limits of the grain size distribution, a_{min} and a_{max} , between 0.001 μm and 5.0 μm ($a_{\text{min}} < a_{\text{max}}$). The dispersion of the normalized data (between days 66 and 239) is added to the error (Fig. 2c). We used the chi-square (χ^2) minimization method to determine the best-fitting parameters α and a_{max} of the grain-size distribution function, for fixed $a_{\text{min}} = 0.001 \mu\text{m}$ (Fig. 3). The χ^2 value for a desired confidence limit p is calculated as $\chi^2 = \chi^2_{\text{min}} + \Delta\chi^2(p)$, where χ^2_{min} is the global minimum χ^2 value of all models. The best-fitting model is characterized by $\alpha = 3.6$, $a_{\text{min}} = 0.001 \mu\text{m}$ and $a_{\text{max}} = 4.2 \mu\text{m}$. However, our models cannot account for the upturn towards H δ (Fig. 2b), which we attribute to a systematic effect caused by intrinsic line changes rather than to small grains. We note that the considered grain radius is truncated at 5 μm , beyond which the size parameter $x = 2\pi a/\lambda$ becomes prohibitively large, making mass absorption coefficient calculations difficult.

Light curves. In Extended Data Fig. 7a we show synthetic UVBRI optical and JHK NIR light curves generated from our X-shooter spectra compared to broad-band photometry from the literature¹² (we have added 1.4 mag to the published U-band magnitudes, which happens to be twice the U-band AB offset). It is evident that there is good agreement between them, giving credence to the flux calibration of our spectra. The energy input (Extended Data Fig. 7b) from ^{56}Co was normalized to the total observed luminosity on about day 26, and the ^{44}Ti contribution was calculated assuming a relative Co/Ti yield (by number) of 3×10^{-4} (ref. 41). The ultraviolet and optical (UVO) and NIR luminosities (Extended Data Fig. 7a) are derived from the black body fits to our X-shooter spectra. A power-law approximation to the UVO luminosity shows that it decays as a $t^{-0.4}$ power law.

Dust heating and vaporization. A dust grain of radius a located at a distance R from the supernova will attain an equilibrium temperature T_d , determined by the balance between the rate it is heated by the supernova and its cooling rate by NIR emission. The equation describing this balance is given by

$$\int_0^\infty \pi a^2 Q_{\text{abs}}(\nu, a) \left(\frac{L_\nu(\nu)}{4\pi R^2} \right) d\nu = \int_0^\infty 4\pi a^2 Q_{\text{abs}}(\nu, a) \pi B_\nu(\nu, T_d) d\nu \quad (3)$$

The supernova will vaporize a grain when its temperature exceeds the vaporization temperature. We take $T_{\text{vap, Si}} = 1,500 \text{ K}$ for silicates. Owing to the large uncertainty in $T_{\text{vap, AC}}$ for amorphous carbon grains, we adopt a temperature range of 2,000–3,000 K.

The supernova light is preceded by a short ($\Delta t \leq 1 \text{ d}$) burst of radiation as the shock, resulting from the core collapse, breaks out of the stellar surface⁴². From direct observations^{43,44} of such a shock breakout and models to fit early ultraviolet optical light curves^{45,46}, a shock breakout burst typically lasts around 100 s to 1,000 s with peak luminosities of around $10^{12} L_\odot$ (where L_\odot is the solar luminosity) and effective temperatures of a few times 10^5 K , after which the luminosities decrease in less than one day by over a few orders of magnitudes. A short burst of radiation, characterized by a 10^5 K black body and a luminosity of $10^{11} L_\odot$ similar to that inferred for Cas A (ref. 47), will vaporize any pre-existing dust in the circumstellar material, creating a dust-free ‘cavity’ of radius R_{cav} . Extended Data Fig. 8a shows that, independently of the grain species, small grains are vaporized out to larger distances from the supernova than large grains. For silicate grains, R_{cav} is about twice as large as for carbon grains. As a consequence of shock breakout, no dust grains can exist out to R_{cav} of about 10^{17} cm to 10^{18} cm .

Any dust that may subsequently form within the radius R_{cav} will be subjected to the supernova light, characterized by a 7,300-K black body and a luminosity of about $5 \times 10^9 L_\odot$. Extended Data Fig. 8b shows the vaporization radii, R_{vap} , for the observed supernova luminosity at the first epoch (26 days past peak). Independently of grain size, the R_{vap} for silicates is significantly larger than R_{vap} for amorphous carbon at any assumed $T_{\text{vap, AC}}$ and CDS radius, which is about $2 \times 10^{16} \text{ cm}$ at this epoch (Supplementary Information, Extended Data Fig. 9b). Dust grains of sizes around 0.05–0.1 μm have the largest vaporization radii for either dust species. It is evident that only amorphous carbon grains can survive the radiation from the underlying supernova at the location of the CDS. Amorphous carbon grains with grain sizes $\geq 0.25 \mu\text{m}$ have temperatures ($\leq 2,200 \text{ K}$) consistent with the hot dust temperatures inferred from the modified black-body fits to the NIR emission (Extended Data Fig. 8c). These grain radii are consistent with those inferred from our extinction measurements. The small carbon grains ($\leq 0.25 \mu\text{m}$), which are required to explain the observed ultraviolet extinction (see Fig. 2), have higher temperatures

(2,200–2,700 K) but do not contribute strongly to the NIR emission. Silicate grains cannot exist at the location of the CDS.

Dust mass estimates. The dust mass is derived from either the extinction or the NIR emission as

$$M_d = N_d \int_{a_{\text{min}}}^{a_{\text{max}}} f(a) m_d(a) da \quad (4)$$

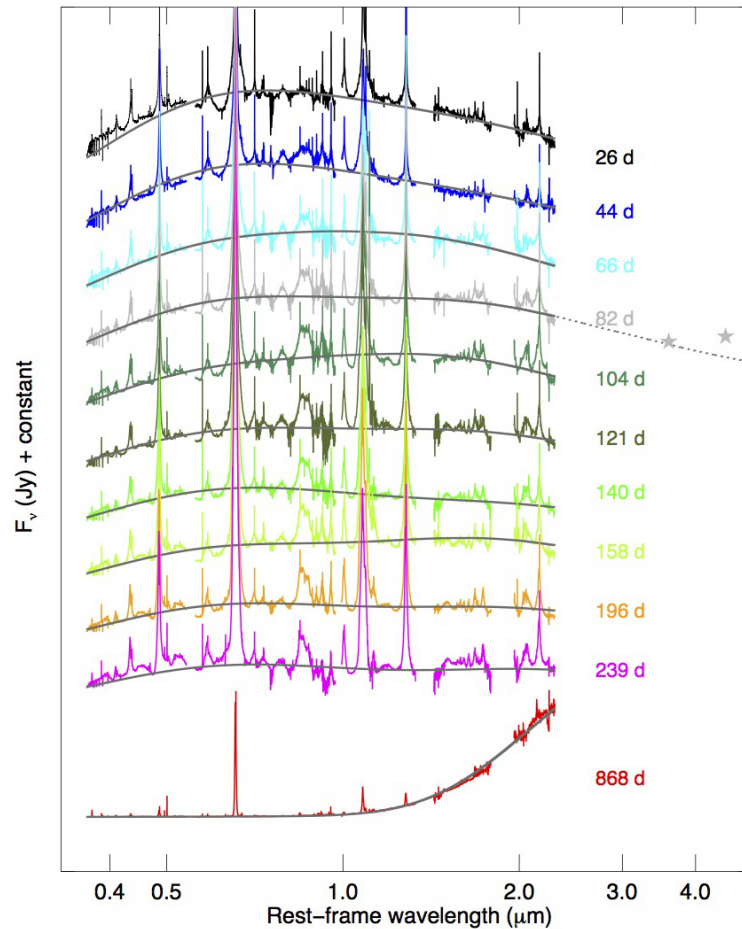
The total number of dust particles N_d is obtained from the fits using either equations (1) or (2).

In Fig. 4 we display the evolution of the carbon dust masses, which are derived from (1) the extinction and its standard deviation obtained for the best-fitting grain size distribution (Figs 2 and 3) at $R_{\text{CDS}} = 2.0 \times 10^{16} \text{ cm}$, and (2) the NIR emission, varying α between 3.5 and 3.7. A power-law fit, $M_d \propto t^\beta$, to the dust mass evolution at early and late phases shows a slow increase ($\beta = 0.8$) at early times and accelerated build-up of the dust mass ($\beta = 2.4$) after 239 days. The estimated carbon dust mass of $\sim 2 \times 10^{-3} M_\odot$ at the late epoch is probably a lower limit (Supplementary Information).

Extended Data Fig. 9a visualizes the sensitivity of the inferred extinction dust mass to a_{max} and α . Carbon dust masses at 239 days past peak are calculated and displayed for parameters within the 3σ confidence interval (see Fig. 3) and for fixed $a_{\text{min}} = 0.001 \mu\text{m}$. Large grains exhibit a stronger dependency on α , with larger dust masses being reached for small α (that is, favouring large grains). For small grains the dust mass is almost independent of α . For large α the dust mass remains independent of a_{max} for large grains, whereas for small α the dust mass increases steeply with increasing a_{max} .

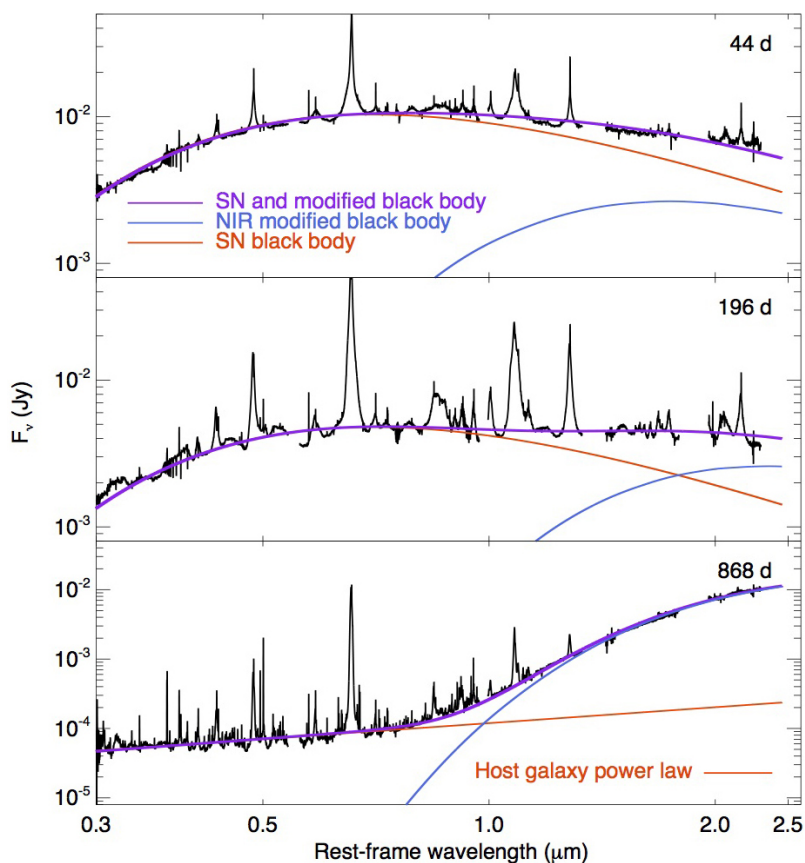
Requiring that the extinction and emission dust mass originate from the CDS, the allowed location of the CDS is constrained by $R_{\text{vap}} \leq R_{\text{CDS}} \leq R_{\text{shock}}$ (Extended Data Fig. 9b). The location of the forward shock R_{shock} at day 239 is estimated assuming a velocity of $3.5 \times 10^4 \text{ km s}^{-1}$ until 26 days past peak and $3,000 \text{ km s}^{-1}$ for the subsequent 213 days.

- Benetti, S. *et al.* Supernova 2010jl in UGC 5189A. *Centr. Bur. Electron. Teleg.* **2536**, 1 (2010).
- Stoll, R. *et al.* SN 2010jl in UGC 5189: yet another luminous type IIn supernova in a metal-poor galaxy. *Astrophys. J.* **730**, 34 (2011).
- D’Odorico, S. *et al.* X-shooter UV- to K-band intermediate-resolution high-efficiency spectrograph for the VLT: status report at the final design review. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 6269* (SPIE, 2006).
- Vernet, J. *et al.* X-shooter, the new wide band intermediate resolution spectrograph at the ESO Very Large Telescope. *Astron. Astrophys.* **536**, A105 (2011).
- Modigliani, A. *et al.* The X-shooter pipeline. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 7737* (SPIE, 2010).
- Horne, K. An optimal extraction algorithm for CCD spectroscopy. *Publ. Astron. Soc. Pacif.* **98**, 609–617 (1986).
- Schlegel, D. J., Finkbeiner, D. P. & Davis, M. Maps of dust infrared emission for use in estimation of reddening and cosmic microwave background radiation foregrounds. *Astrophys. J.* **500**, 525–553 (1998).
- Rouleau, F. & Martin, P. G. Shape and clustering effects on the optical properties of amorphous carbon. *Astrophys. J.* **377**, 526–540 (1991).
- Li, A. & Draine, B. T. Infrared emission from interstellar dust. II. The diffuse interstellar medium. *Astrophys. J.* **554**, 778–802 (2001).
- Fransson, C. *et al.* High density circumstellar interaction in the luminous type IIn SN 2010jl: the first 1100 days. Preprint at <http://arxiv.org/abs/1312.6617> (2013).
- Rauscher, T., Heger, A., Hoffman, R. D. & Woosley, S. E. Nucleosynthesis in massive stars with improved nuclear and stellar physics. *Astrophys. J.* **576**, 323–348 (2002).
- Dwek, E. The infrared echo of a type II supernova with a circumstellar dust shell—applications to SN 1979c and SN 1980k. *Astrophys. J.* **274**, 175–183 (1983).
- Soderberg, A. M. *et al.* An extremely luminous X-ray outburst at the birth of a supernova. *Nature* **453**, 469–474 (2008).
- Malesani, D. *et al.* Early spectroscopic identification of SN 2008D. *Astrophys. J.* **692**, L84–L87 (2009).
- Blinnikov, S., Lundqvist, P., Bartunov, O., Nomoto, K. & Iwamoto, K. Radiation hydrodynamics of SN 1987A. I. Global analysis of the light curve for the first 4 months. *Astrophys. J.* **532**, 1132–1149 (2000).
- Tolstov, A. G., Blinnikov, S. I. & Nadyozhin, D. K. Coupling of matter and radiation at supernova shock breakout. *Mon. Not. R. Astron. Soc.* **429**, 3181–3199 (2013).
- Dwek, E. & Arendt, R. G. Infrared echoes reveal the shock breakout of the Cas A supernova. *Astrophys. J.* **685**, 976–987 (2008).



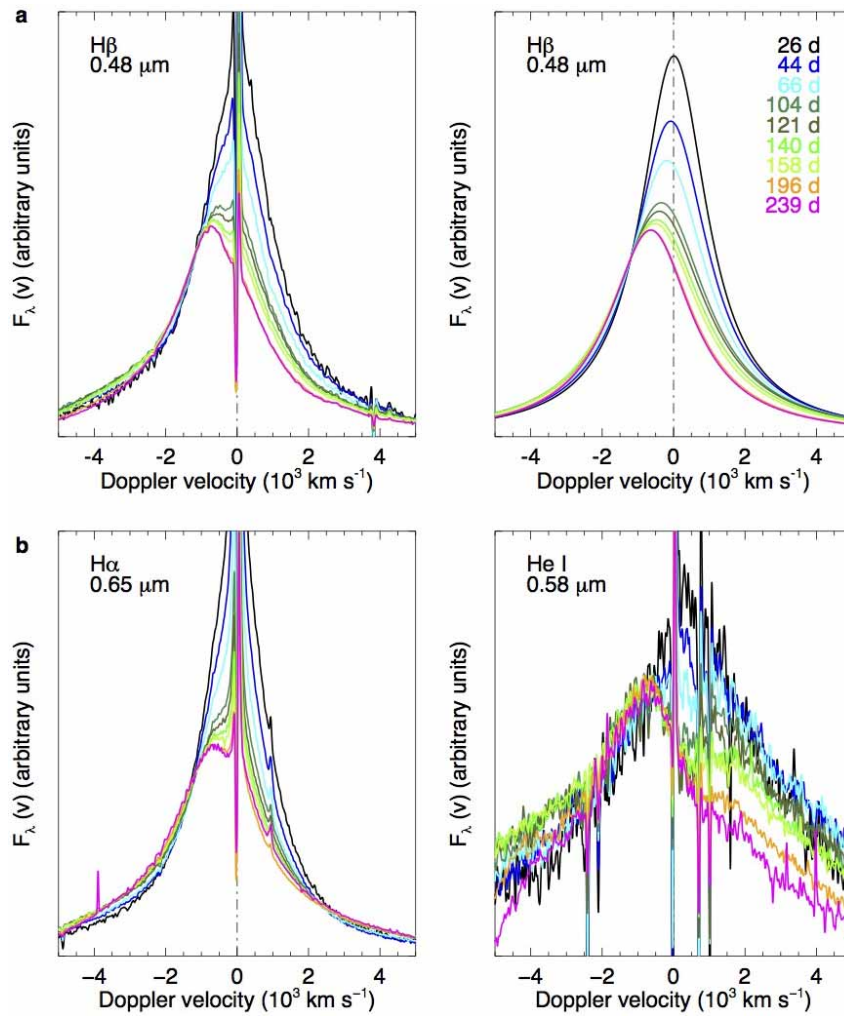
Extended Data Figure 1 | Time sequence of the supernova spectra. Spectra (flux density (Jy)) from ten epochs between $t = 26$ days and 868 days past peak. The spectra are offset by an arbitrary constant. The atmospheric telluric bands at 1.33–1.43 μm and 1.79–1.96 μm have been excluded, as well as the dichroic gaps between the X-shooter instrument arms at 0.54–0.56 μm and

0.97–0.995 μm . The light-grey spectrum is an interpolated spectrum at the epoch of observations of the Infrared Array Camera 3.6 μm and 4.5 μm data (grey stars)¹¹. The solid grey curves are fits to the spectra, composed of multiple distinct black-body functions.



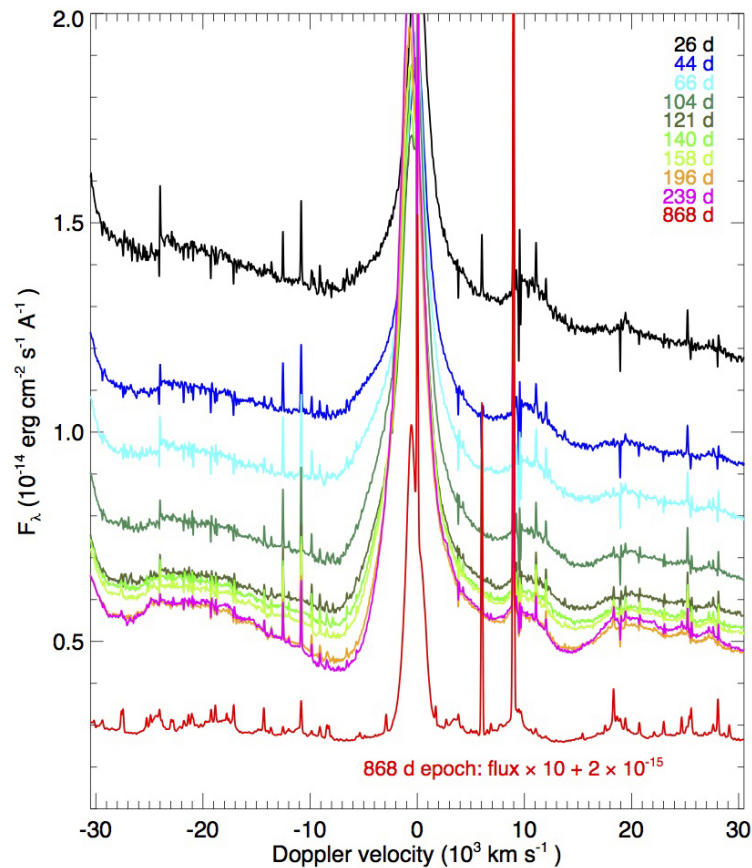
Extended Data Figure 2 | NIR excess dust emission in supernova spectra at three different epochs. The spectral shape of the supernova (SN) shows little evolution for the early epochs (44 and 196 days past peak). The late epoch at 868 days exhibits strong NIR emission while the supernova continuum has

faded. The atmospheric telluric bands at 1.33–1.43 μm and 1.79–1.96 μm , as well as the dichroic gaps of the X-shooter instrument arms at 0.54–0.56 μm and 0.97–0.995 μm , have been excluded.



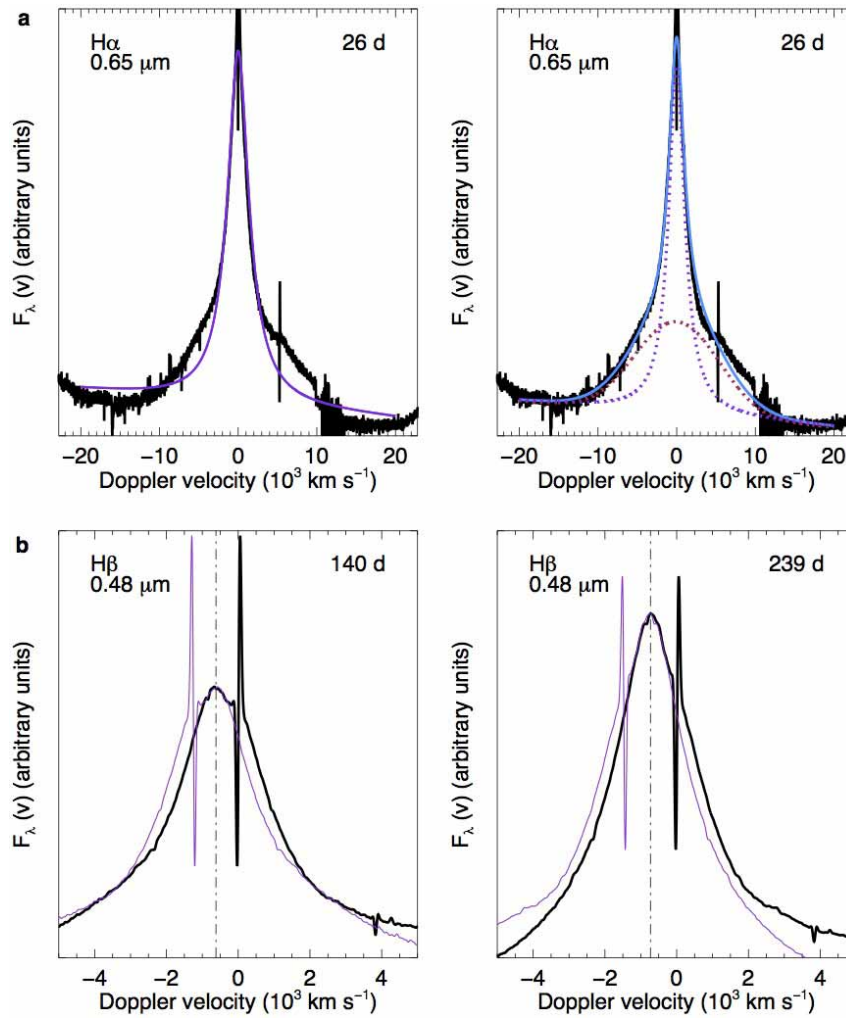
Extended Data Figure 3 | Line profiles. **a**, Comparison of the observed line profile (left panel) to the line profile of the Lorentzian line fits (right panel), illustrated for H β λ 4,861.35. **b**, The left panel shows the line profile of the H α λ 6,562.79 line. The progressive broadening of the line causes both the blue and

red wings to cross at different epochs. The right panel shows the line profile of the He I λ 5,875.621 line exhibiting a similar effect. The lines increasingly deviate from a Lorentzian profile.



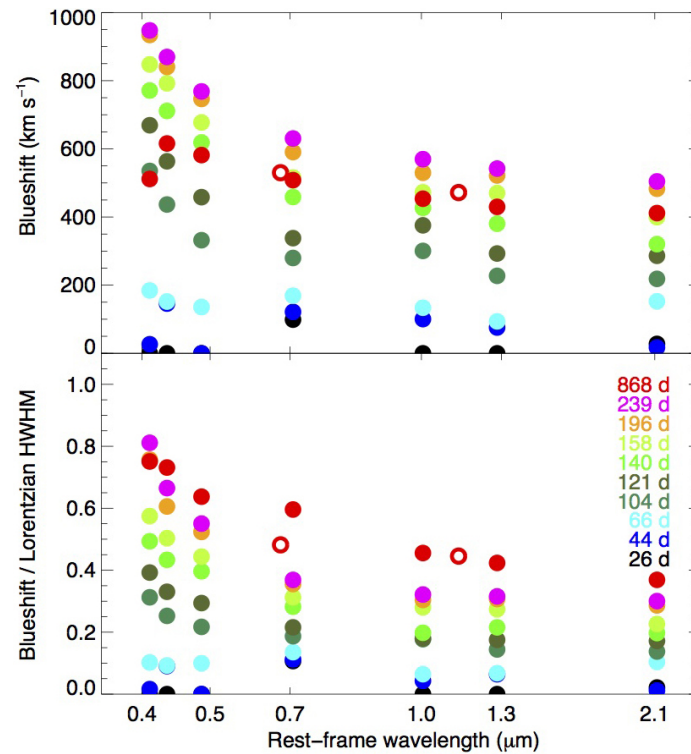
Extended Data Figure 4 | Development of the broad P Cygni profile of H β . Within the early epochs (<239 days) the hydrogen emission line H β λ 4,861.35 develops a strong P Cygni profile. The minimum of the P Cygni profile is at about $7,500 \text{ km s}^{-1}$. The largest velocities associated with the P Cygni profile

are at about $20,000 \text{ km s}^{-1}$. The late epoch (868 days) has been scaled by a factor of ten and offset for better comparison to the early epochs. The H β line no longer exhibits features of high velocities. The wings of the intermediate-velocity component extend to around $2,000\text{--}3,000 \text{ km s}^{-1}$.



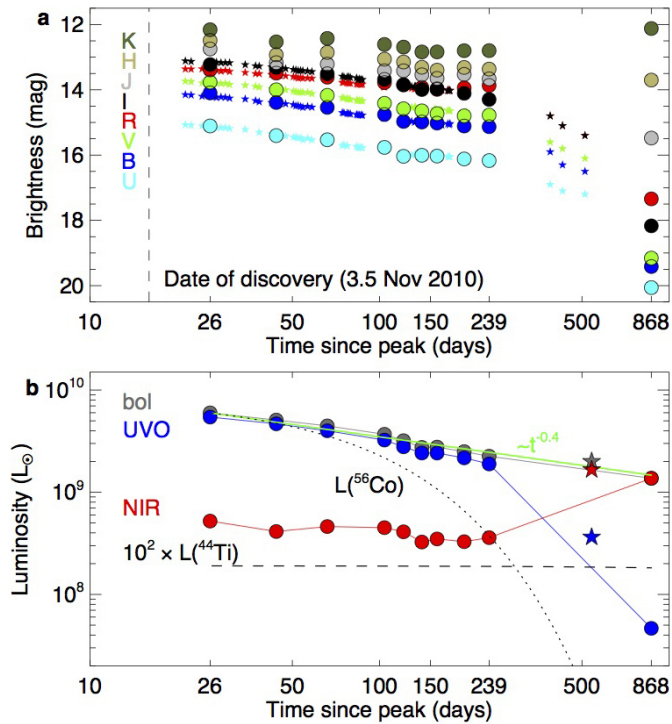
Extended Data Figure 5 | Velocity components and asymmetry of the intermediate emission lines. **a**, The left panel shows that the $H\alpha$ $\lambda 6,562.79$ line cannot be fitted with a single Lorentzian (purple solid curve). The right panel shows the broad (pink dotted curve) and the intermediate-velocity component (purple dotted curve) and the combination of the two (blue solid

curve). **b**, The $H\beta$ $\lambda 4,861.35$ line is asymmetric with respect to its peak velocities (approximately -458 km s^{-1} at 140 days and approximately -768 km s^{-1} at 239 days). The mirrored emission lines are shown as thin purple curves. The mirror axis is shown as a black dashed-dotted curve. Similar effects are seen for other emission lines.

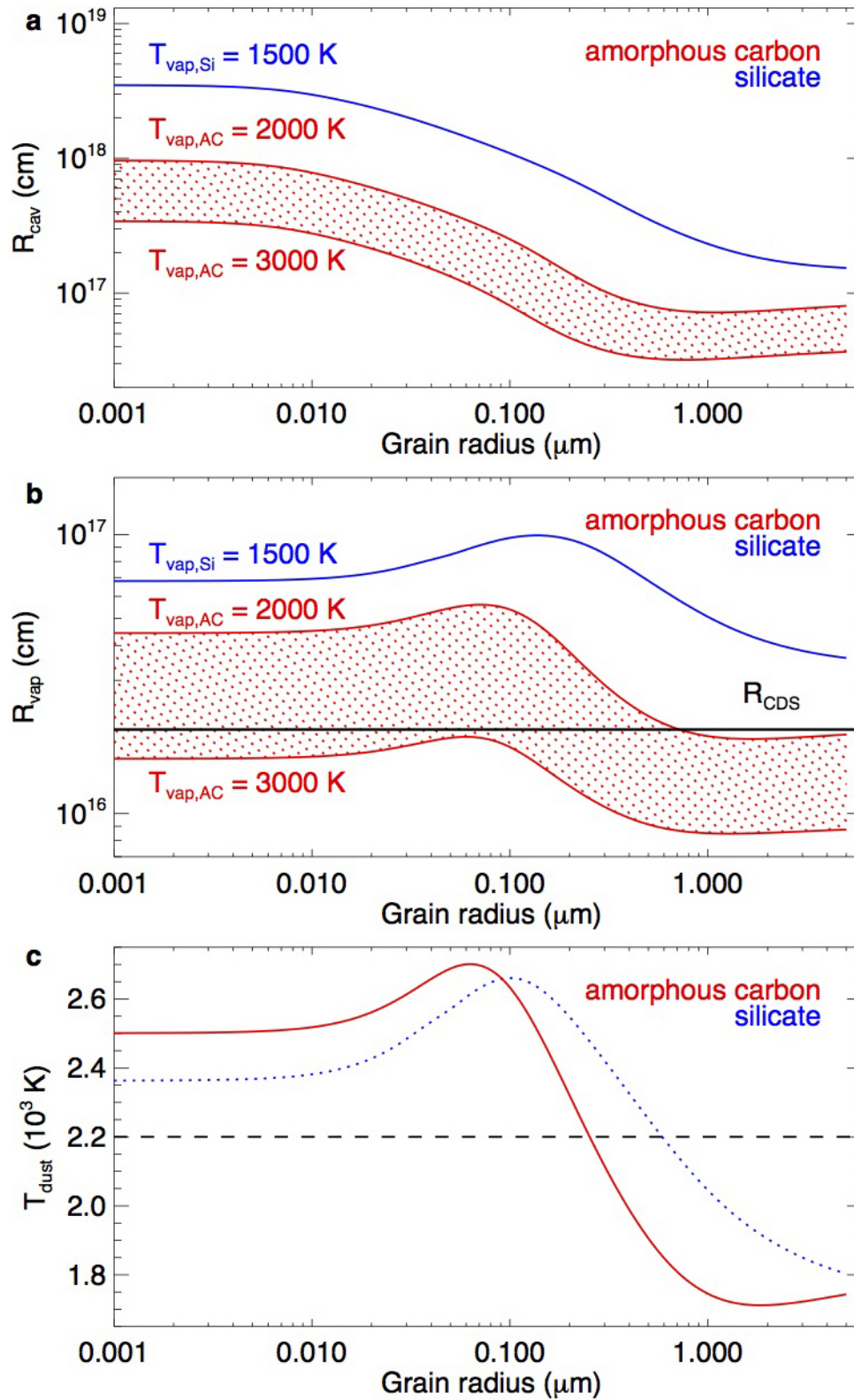


Extended Data Figure 6 | Evolution of the blueshift velocity of hydrogen and metal lines. The blueshift of the hydrogen lines is wavelength-dependent and increases with time for the early epochs. At any epoch the blueshift is smaller for lines at longer wavelengths. The filled symbols correspond to the

blueshifts of the hydrogen emission lines and the open circles correspond to the oxygen lines. The blueshift-to-HWHM ratio for the early epochs resembles the extinction curves (Fig. 2).

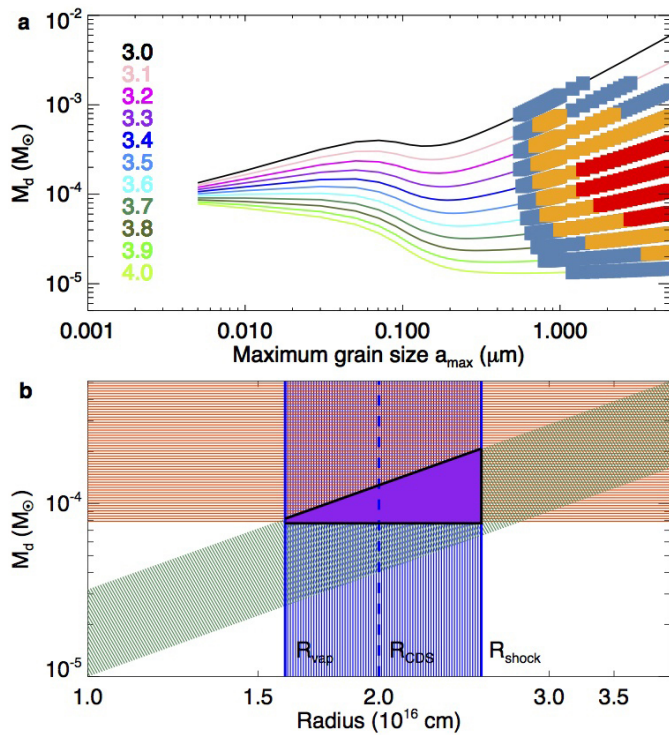


Extended Data Figure 7 | Light curves. **a**, Synthetic UVBRI and JHK light curves (filled circles) compared to the UBVR optical photometry of ref. 12 (small stars). **b**, Energy output. The temporal evolution of the UVO and NIR luminosities (blue and red symbols, respectively) and the total bolometric (UVO + NIR) luminosity (black diamonds). The green curve is a $t^{-0.4}$ power-law approximation to the UVO emission at early times. We have included data points from the literature (filled stars) at 553 days (ref. 12). The maximum possible contributions to the heating of the ejecta from the radioactively decaying ^{56}Co and the isotope ^{44}Ti are shown as a dotted curve and a dashed line, respectively.



Extended Data Figure 8 | Dust vaporization radii and temperatures as a function of grain radius. **a**, Radii R_{cav} from an initial burst of radiation. **b**, Radii R_{vap} from the observed supernova luminosity at 26 days. R_{cav} and R_{vap} depend on the vaporization temperatures $T_{\text{vap,AC}}$ and $T_{\text{vap,Si}}$. The black line indicates the location, R_{CDS} , of the CDS. **c**, The dust temperatures at R_{CDS} , for

grains heated by the supernova light and cooled through the NIR emission. The dashed line indicates T_{hot} derived from the spectral fits (26 days). Amorphous carbon grains (solid curve) have temperatures $\leq T_{\text{vap,AC}}$. Silicate grains (dotted curve) would be hotter than $T_{\text{vap,Si}}$ and therefore cannot exist.



Extended Data Figure 9 | Dust mass at 239 days past peak. **a**, Sensitivity of the dust mass to the parameters a_{max} (coloured curves) and α of the grain-size distribution function. The filled coloured squares represent the dust masses for the parameters of the grain size distribution function of the 1 σ (red), 2 σ (orange) and 3 σ (blue) confidence intervals (Figs 2c and 3). **b**, The extinction dust mass and its standard deviation (green-shaded band), the dust mass from the NIR emission (red-shaded band) and the radius range $R_{\text{vap}} \leq R_{\text{CDS}} \leq R_{\text{shock}}$ (blue lines and shaded area). The overlapping region (purple framed area) of the three bands constrains the radius of the CDS (R_{CDS}) and the dust mass.

Extended Data Table 1 | Log of the VLT/X-shooter observations of SN 2010jl

Date (UT)	Airmass	Seeing (")	Exposure times (s)			Days past peak on 2010 Oct 18.6 UT
			UVB	VIS	NIR	
2010 Nov 13.4	1.53	0.91	2×100	2×100	2×100	26
2010 Dec 1.4	1.25	1.74	2×250	2×250	8×100	44
2010 Dec 23.3	1.22	1.52	2×250	2×250	8×100	66
2011 Jan 30.3	1.21	1.05	2×250	2×250	8×100	104
2011 Feb 16.1	1.34	0.87	2×250	2×250	8×100	121
2011 Mar 7.2	1.24	0.77	2×250	2×250	8×100	140
2011 Mar 25.1	1.21	0.98	2×400	2×450	10×100	158
2011 May 3.0	1.21	0.74	2×400	2×450	10×100	196
2011 Jun 15.0	1.81	0.90	4×550	4×600	32×100	239
2013 Mar 4.0	1.28	0.86	8×698	8×605	56×100	868

Ramp compression of diamond to five terapascals

R. F. Smith¹, J. H. Eggert¹, R. Jeanloz², T. S. Duffy³, D. G. Braun¹, J. R. Patterson¹, R. E. Rudd¹, J. Biener¹, A. E. Lazicki¹, A. V. Hamza¹, J. Wang², T. Braun¹, L. X. Benedict¹, P. M. Celliers¹ & G. W. Collins¹

The recent discovery of more than a thousand planets outside our Solar System^{1,2}, together with the significant push to achieve inertially confined fusion in the laboratory³, has prompted a renewed interest in how dense matter behaves at millions to billions of atmospheres of pressure. The theoretical description of such electron-degenerate matter has matured since the early quantum statistical model of Thomas and Fermi^{4–10}, and now suggests that new complexities can emerge at pressures where core electrons (not only valence electrons) influence the structure and bonding of matter¹¹. Recent developments in shock-free dynamic (ramp) compression now allow laboratory access to this dense matter regime. Here we describe ramp-compression measurements for diamond, achieving 3.7-fold compression at a peak pressure of 5 terapascals (equivalent to 50 million atmospheres). These equation-of-state data can now be compared to first-principles density functional calculations¹² and theories long used to describe matter present in the interiors of giant planets, in stars, and in inertial-confinement fusion experiments. Our data also provide new constraints on mass-radius relationships for carbon-rich planets.

Mass-radius data for extrasolar planets combined with equation-of-state (EOS) models for constituent materials reveal that matter at pressures of several terapascals is quite common throughout the Universe^{1,2,13}. At several terapascals, matter is approaching an atomic-scale pressure (for example, the quantum-mechanical ‘pressure’ that counteracts the electrons’ Coulomb attraction in a Bohr atom), at which material structure and chemistry, and even the properties of atoms themselves, are expected to change¹¹. Recent density functional theory (DFT) calculations predict that in several materials electrons become localized at terapascal conditions, with structural and electronic complexity unexpected from quantum statistical models (such as that of Thomas and Fermi)¹².

Experimental access to multi-terapascal conditions is now possible with dynamic ramped compression. Dynamic compression is necessary to achieve atomic-scale pressures, conditions far beyond those accessible in static experiments¹⁴. Ramp compression produces less dissipative heating, thus enabling higher compression and lower temperature than does shock compression¹⁵. However, ramp compression is unstable relative to a shock because sound velocities typically increase with pressure, so precise control of the applied pressure-loading history is required to achieve high pressures without shock formation.

The National Ignition Facility, a 2-MJ laser designed to create thermonuclear fusion in the laboratory³, offers the energy and control necessary to ramp compress matter to several terapascals. Here we describe ramp-loading measurements on carbon to 5 TPa, with stress, density and sound speed determined for the entire compression path. These unprecedented conditions provide experimental constraints on the carbon EOS at pressures more than thirty times that of previous static-compression measurements, and where state-of-the-art DFT coincides with modern versions of the quantum-statistical Thomas–Fermi model, originally developed early in the past century to describe matter at extreme compressions.

In these experiments, 176 laser beams deliver a total peak power of 2.2 TW, with accuracy of better than 1% in power and 0.02 ns in time, over a duration of 20 ns. The light hitting a target (indirectly) creates an

ablatively driven pressure wave in the sample (Fig. 1), and—because pressure scales as the 7/8th power of the laser intensity¹⁶—the pressure is controlled to better than 1%. Samples consist of nanocrystalline diamond, shaped with steps so that the pressure-wave transit across four different thicknesses is recorded for each experiment. Response of the sample is characterized by velocity interferometry (VISAR), which records the velocity of the sample’s free (back) surface as it is engulfed by the pressure wave (Fig. 1). Iterative Lagrangian analysis is used to translate these velocity data into a stress–density relation that quantifies the loading path (Fig. 2)¹⁷. These data are absolute—not referenced against a

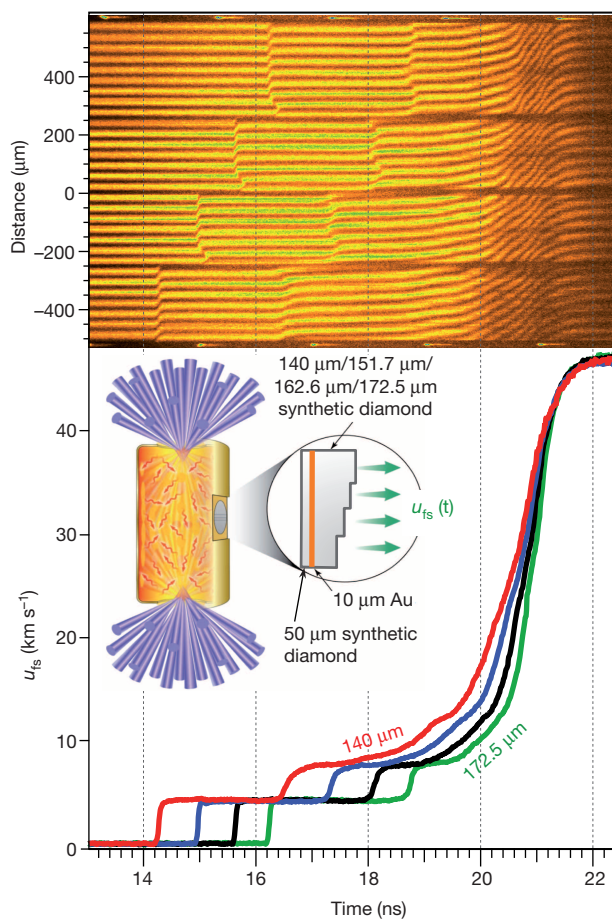


Figure 1 | Velocity interferometry for ramp compressed diamond. Top, the temporally resolved velocity interferometry record. Bottom, derived free-surface velocity u_{fs} versus time. The target (inset) consists of a gold cylinder (hohlraum) 6 mm in diameter by 11 mm long, inside which the 351-nm-wavelength laser light (purple beams) is converted to X-ray energy that is absorbed by the diamond sample attached to the side of the hohlraum. The X-rays ablate and ramp-compress the sample, and the free-surface velocity is recorded for four thicknesses of diamond: 140.0 μm (red line), 151.7 μm (blue line), 162.6 μm (black line) and 172.5 μm (green line) (see Methods).

¹Lawrence Livermore National Laboratory, PO Box 808, Livermore, California 94550, USA. ²Department of Earth and Planetary Science, Department of Astronomy and Miller Institute for Basic Research in Science, University of California, Berkeley, California 94720, USA. ³Department of Geosciences, Princeton University, Princeton, New Jersey 08544, USA.

standard—which is important for quantifying the EOS and benchmarking condensed-matter theories in the terapascal regime.

In detail, we initiate loading with a shock wave of approximately 0.1 TPa, before the onset of the main ramp compression (Fig. 1). Such pre-ramp loading of diamond produces a more fluid-like (strength-free) state¹⁸, which is important for reducing the dissipative heating that can limit compression. Longitudinal stress (P_x)—not pressure—is shown in Fig. 2, because our one-dimensional loading method creates a uniaxial strain that relaxes towards an isotropic state.

A typical record (Fig. 1) shows a free-surface velocity profile $u_{fs}(t)$, characterized by an initial shock to 4.1 km s^{−1}, followed by a fast rise and plateau at 7.2 km s^{−1}, and subsequent ramp compression to 46.6 km s^{−1} (3.7 TPa). Our analysis yields the Lagrangian sound speed (C_L) and P_x as functions of density ρ from the measured $u_{fs}(t)$ (Fig. 2)¹⁷. In all, three experiments yielded $C_L(\rho)$ and $P_x(\rho)$ to peak stresses of 2.7 TPa, 3.7 TPa and 5 TPa, respectively. C_L decreases abruptly at $u_{fs} = 4.1$ km s^{−1}, corresponding to a longitudinal stress of $P_{x,limit} = 0.11$ TPa, which we interpret to be the dynamic strength (elastic limit) of diamond. This also shows

up as the slight deviation in the stress–density relation near 0.11 TPa (Fig. 2, inset). Hydrodynamic simulations indicate that the rapid rise and plateau in $u_{fs}(t)$ at 7.2 km s^{−1} corresponds to a reverberating compression wave within the intermediate Au layer (Fig. 1).

These new data are compared to several carbon EOS models in the multi-terapascal regime (Fig. 2, Extended Data Fig. 1, Extended Data Table 1, and Methods). A cold curve derived from first-principles DFT¹² is in good agreement with a Mie–Grüneisen reduction and extrapolation of shock-Hugoniot data collected to 2 TPa. Also shown are the cold curve formulations from Vinet¹⁹ and Birch–Murnaghan²⁰ each fitted to existing diamond anvil cell data^{21,22}. (Even at these extreme pressures, the differences between the room-temperature isentrope and isotherm and the cold curve (0 K) are indistinguishable on this scale, so for consistency, we refer below simply to the cold curve.) For reference, the Hugoniots calculated from both DFT (solid red line) and a Mie–Grüneisen model (solid orange line) are shown in Fig. 2b. The DFT Hugoniot predicts carbon to be liquid and much less compressible than the DFT cold curve for stresses above about 1 TPa. The differences between the cold curves (grey band) and Hugoniots (orange band) in Fig. 2b illustrate the uncertainties in using prior data for extrapolating the carbon EOS into the terapascal regime.

The cold curve calculated by DFT shows a sequence of phase transformations: diamond to BC8 (body-centred cubic $1a\bar{3}$) (at ~ 0.99 TPa), BC8 to simple cubic (at ~ 2.7 TPa)¹², which are apparent in stress–density curves as stress plateaus corresponding to increased densities (Fig. 2b). No such stress plateaus are apparent in our data. Although phase-transformation kinetics can smooth such features²³, determining whether or not these phase transformations occur will require further work²⁴. Metadynamics calculations for carbon do indicate that the diamond-to-BC8 transition kinetics may be quite slow²⁵.

Static compression and elasticity measurements^{21,22} up to their highest pressures (0.15 TPa) are indistinguishable from the DFT cold curve and standard EOS model fits to the data (Vinet and Birch–Murnaghan). However, when extrapolated to 5 TPa these models differ by about 20% in density (Fig. 2 and Fig. 3, inset). Our data lie between these cold curve calculations.

Also consistent with the DFT cold curve are the gradient-corrected (TFD-W) and the gradient-and-correlation-corrected (TFD-Wc) Thomas–Fermi–Dirac EOSs between about 2 TPa and 5 TPa (Fig. 2)⁹. This agreement is notable because the statistical-atom model considers neither crystal structure nor orbital information, whereas DFT includes both. This agreement may be partly fortuitous, because carbon might not yet be in its densest crystal structure at these pressures, and the deviation of statistical-atom theories is towards predicting densities that are systematically too low.

Our ramp data achieve higher density than the shock Hugoniot, consistent with temperatures being lower for ramp compression versus shock compression^{15,26}. Moreover, these new data are somewhat less compressible than cold-isothermal compression calculations with DFT over most of the pressure range studied, and modern Thomas–Fermi–Dirac formulations (TFD-W and TFD-Wc). We expect that the overlap of the ramp compression data with the older uncorrected Thomas–Fermi–Dirac data in the 2–3 TPa regime is fortuitous. Sample temperature, material strength¹⁸ and phase transformation kinetics²³ can each cause a less compressible stress–density path with respect to the cold curve, so these data should be considered an upper bound for such comparison. Indeed, further study is needed to obtain a better understanding of the differences between theory and experiment and to develop measurement techniques (such as for temperature and structural determination) with which to explore this new extreme matter regime.

The experimental techniques developed here provide a new capability to experimentally reproduce pressure–temperature conditions deep in planetary interiors. Carbon is the fourth most abundant element in the cosmos and has a potentially important role in many types of planets, both within and outside the Solar System. One proposed group of super-Earth exoplanets (1–10 Earth masses in size) are those enriched in carbon,

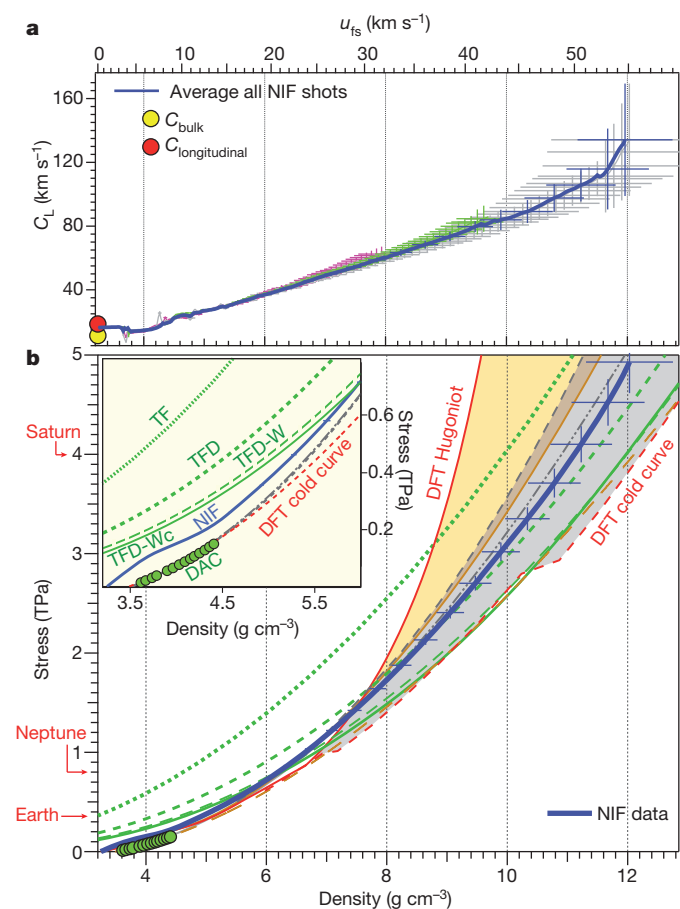


Figure 2 | Ramp compression stress and sound velocity measurements. **a**, Lagrangian sound velocity C_L versus density. **b**, Longitudinal stress P_x versus density. Three experiments (pink, light-green and grey lines) yield C_L data and their average (dark blue line), which are used to determine P_x –density¹⁷ (dark blue line in **b**). Error bars, 1σ . Model comparisons include DFT (solid red line)¹⁰ and Mie–Grüneisen (solid orange line) Hugoniots (density correction discussed in Methods); cold curves from DFT¹² (red dashed line), statistical-atom models (TF, TFD, TFD-W and TFD-Wc as green dotted, short dashed, long dashed and solid lines)⁹, and Vinet¹⁹ (grey dot-dashed line) and Birch–Murnaghan²⁰ (grey dashed line) EOS fits to static data^{21,22}. Pressure-scale-corrected²¹ static diamond anvil cell (DAC) data²² are green circles. Shaded regions between cold curves (grey) or Hugoniots (orange) show roughly the range of uncertainty in the EOS in this terapascal regime. Central pressures for Earth, Neptune and Saturn are shown for reference. The inset highlights the differences in the models at low pressure.

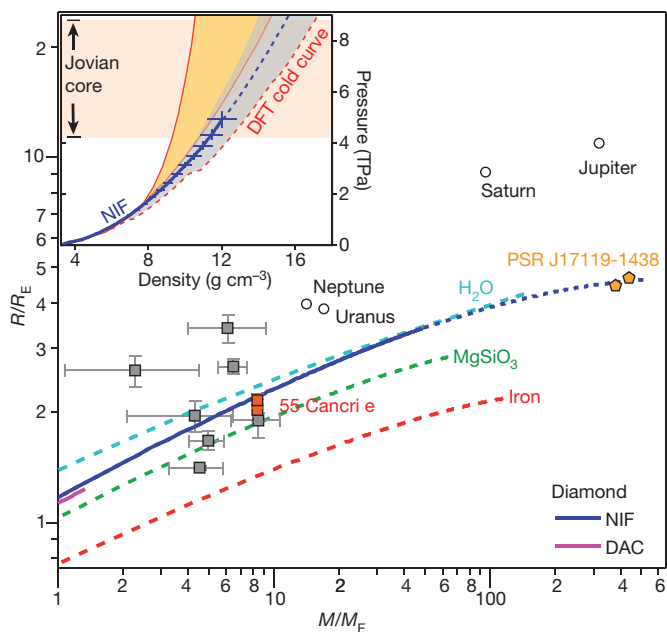


Figure 3 | Mass–radius relationships for homogenous-composition planets. Calculations for carbon (based on our data, where 1σ error bars are within the width of the line, dark blue), H_2O (light blue), post-perovskite MgSiO_3 (green) and iron (red)^{13,30} (lines are dashed when based on extrapolated EOS data). Yellow symbols are values consistent with the minimum density for the companion object to pulsar PSR J1719-1438 for assumed orbital inclinations of 90 and 60 degrees²⁹. The grey squares represent selected transiting super-Earths, with error bars as reported in ref. 27. Two possible values of radii R are shown for 55 Cancri e (red squares)²⁷. The inset shows P –density relevant to Jupiter’s core (~ 4.3 – 8.8 TPa)²⁸ with other curves as in Fig. 2. M_E and R_E are the mass and radius of the Earth, respectively.

and the planet 55 Cancri e has been proposed as a possible carbon planet²⁷. Figure 3 shows mass–radius relationships for selected known super-Earths together with various hypothetical uniform-composition planets, including a pure-carbon planet based on our ramp-compression EOS. Using our new data, we find the central pressure for a 10-Earth-mass pure-carbon planet to be about 0.8 TPa. This new capability to reach multi-terapascal pressures also enables experimental access to Jupiter’s core pressures²⁸ where extrapolations of earlier shock and static data become unreliable (Fig. 3, inset).

Our results also have relevance for large pulsar planets, such as the companion of millisecond pulsar PSR J1719-1438 (ref. 29). This object has a minimum mass somewhat larger than Jupiter (1.15×10^{-3} solar masses or 383 Earth masses), and a 2.2-hour orbital period. A carbon-rich composition was suggested, based on TFD-Wc results for carbon^{9,29}. The reliability of this form of TFD theory as shown by our experiments supports this interpretation. An extrapolation of our EOS is consistent with TFD-Wc in suggesting that an object of this mass made of pure carbon would have a radius of about 4.5 Earth radii and a central pressure of about 148 TPa. The mean density of 23 g cm^{-3} is compatible with the measured minimum density of the pulsar planet²⁹.

In summary, diamond, the least compressible material known, has here been compressed to an unprecedented density of 12 g cm^{-3} , more than that of lead at ambient conditions. The measured Lagrangian sound speed, stress and density provide the first experimental data for constraining condensed-matter theory and planet-evolution models in the terapascal regime. By realizing three necessary conditions—(1) the adiabatic conditions of dynamic compression; (2) a loading profile soft enough to avoid shock formation; and (3) a nearly fluid-like response of the sample such that strength and dissipation are minimal—these experiments document an approach for taking solids to the long-sought high-density conditions of statistical-electron theory.

METHODS SUMMARY

Experiments used 176 laser beams from the National Ignition Facility (NIF) (in Livermore, California, USA) focused onto the inner walls of a gold hohlraum (a gold cylinder that converts the laser light to X-rays) with a combined laser energy up to 0.76 MJ in a ~ 20 -ns temporally ramped pulse. This generates a spatially uniform near-blackbody distribution of thermal X-rays in the hohlraum with a characteristic radiation temperature T_r , which increases with time to a peak of $T_r \approx 235 \text{ eV}$. The subsequent X-ray ablation of the diamond, over a 3-mm diameter, produces a uniform ramp-compression wave, which outruns the thermal wave produced by ablation. As the pressure wave reaches the back surface of the diamond the free surface velocity of each step is recorded with an imaging velocity interferometer (Fig. 1).

Samples consist of a 50- μm -thick diamond plate used as an ablator, a 10- μm Au layer preheat shield, and a diamond sample having four steps (Fig. 1 inset). The diamond was synthesized by chemical vapour deposition to yield a layered microstructure with an average grain size of 200 nm and a density of 3.2491 g cm^{-3} ($\pm 0.01\%$). The final sample had alternating 0.35- μm layers of 20-nm grains and ~ 350 -nm grains. X-ray diffraction showed a $\langle 110 \rangle$ texture in the growth direction. The thickness of the composite sample is determined to $\pm 1.0 \mu\text{m}$, and the differences in step thickness are determined by optical interferometry to $\pm 0.1 \mu\text{m}$. The Au layer was incorporated into the target design to serve as a radiation preheat shield for the step diamond sample. Detailed radiation transport simulations estimate a temperature rise of 33 K due to X-ray preheating.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 October 2013; accepted 16 May 2014.

- Schneider, J., Dedieu, C., Le Sidaner, P., Savalle, R. & Zolotukhin, I. Defining and cataloging exoplanets: the exoplanet.eu database. *Astron. Astrophys.* **532**, A79 (2011).
- Seager, S., Kuchner, M., Hier-Majumder, C. A. & Militzer, B. Mass-radius relationships for solid exoplanets. *Astrophys. J.* **669**, 1279–1297 (2007).
- Edwards, M. J. et al. Progress towards ignition on the National Ignition Facility. *Phys. Plasmas* **20**, 070501 (2013).
- Thomas, L. H. The calculation of atomic fields. *Math. Proc. Camb. Phil. Soc.* **23**, 542–548 (1927).
- Dirac, P. A. M. Note on exchange phenomena in the Thomas atom. *Math. Proc. Camb. Phil. Soc.* **26**, 376–385 (1930).
- Feynman, R. P., Metropolis, N. & Teller, E. Equation of state of elements based on the generalized Fermi-Thomas theory. *Phys. Rev.* **75**, 1561–1573 (1949).
- Salpeter, E. E. & Zapolsky, H. S. Theoretical high pressure equations of state, including correlation energy. *Phys. Rev.* **158**, 876–886 (1967).
- Abrahams, A. M. & Shapiro, S. L. Cold equation of state from Thomas-Fermi-Dirac-Weizsacker theory. *Phys. Rev. A* **42**, 2530–2538 (1990).
- Lai, D., Abrahams, A. M. & Shapiro, S. L. Equation of state in metals and cold stars: evaluation of statistical models. *Astrophys. J.* **377**, 612–628 (1991).
- Correa, A. A., Benedict, L. X., Young, D. A., Schwegler, E. & Bonev, S. A. A first principles multi-phase equation of state of carbon under extreme conditions. *Phys. Rev. B* **78**, 024101 (2008).
- Neaton, J. B. & Ashcroft, N. W. Pairing in dense lithium. *Nature* **400**, 141–144 (1999).
- Martinez-Canales, M., Pickard, C. J. & Needs, R. J. Thermodynamically stable phase of carbon at multi-terapascal pressures. *Phys. Rev. Lett.* **108**, 045704 (2012).
- Swift, D. C. et al. Mass-radius relationships for exoplanets. *Astrophys. J.* **744**, 59–68 (2012).
- Dubrovinsky, L., Dubrovinskaya, N., Prakapenka, V. B. & Abakumov, A. M. Implementation of micro-ball nanodiamond anvils for high-pressure studies above 6 Mbar. *Nature Commun.* **3**, 1163 (2012).
- Zel'dovich, Ya. B. & Raizer, Yu. P. *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena* (Dover, 2002).
- Atzeni, S. & Meyer-ter-Vehn, J. *The Physics of Inertial Fusion: Beam Plasma Interaction, Hydrodynamics, Hot Dense Matter* (Oxford Univ. Press, 2004).
- Rothman, S. D. et al. Measurement of the principle isentropes of lead and lead-antimony alloy to ~ 400 kbar by quasi-isentropic compression. *J. Phys. D* **38**, 733–740 (2005).
- McWilliams, R. S. et al. Strength effects in diamond under shock compression from 0.1 to 1 TPa. *Phys. Rev. B* **81**, 014111 (2010).
- Vinet, P., Ferrante, J., Rose, J. H. & Smith, J. R. Compressibility of solids. *J. Geophys. Res.* **92**, 9319–9325 (1987).
- Birch, F. Finite elastic strain of cubic crystals. *Phys. Rev.* **71**, 809–824 (1947).
- Dewaele, A., Datchi, F., Loubeyre, P. & Mezouar, M. High pressure-high temperature equation of state of neon and diamond. *Phys. Rev. B* **77**, 094106 (2008).
- Occelli, F., Loubeyre, P. & Letoullec, R. Properties of diamond under hydrostatic pressures up to 140 GPa. *Nature Mater.* **2**, 151–154 (2003).
- Smith, R. F. et al. Time-dependence of the alpha to epsilon phase transformation in iron. *J. Appl. Phys.* **114**, 223507 (2013).
- Coppari, F. et al. Experimental evidence for a phase transition in magnesium oxide at exoplanet pressures. *Nature Geosci.* **6**, 926–929 (2013).

25. Sun, J., Klug, D. D. & Martoňák, R. Structural transformations in carbon under extreme pressure: beyond diamond. *J. Chem. Phys.* **130**, 194512 (2009).
26. Eggert, J. H. *et al.* Melting temperature of diamond at ultrahigh pressure. *Nature Phys.* **6**, 40–43 (2010).
27. Madhusudhan, N., Lee, K. K. M. & Mousis, O. A possible carbon-rich interior in super-Earth 55 Cancri e. *Astrophys. J.* **759**, L40 (2012).
28. Nettelmann, N. *et al.* Ab initio equation of state data for hydrogen, helium, and water and the internal structure of Jupiter. *Astrophys. J.* **683**, 1217–1228 (2008).
29. Bailes, M. *et al.* Transformation of a star into a planet in the millisecond Pulsar binary. *Science* **333**, 1717–1720 (2011).
30. Wagner, F. W., Sohl, F., Hussmann, H., Grott, M. & Rauer, H. Interior structure models of solid exoplanets using material laws in the infinite pressure limit. *Icarus* **214**, 366–376 (2011).

Acknowledgements We thank the NIF staff, B. Goldstein, Ed Moses, C. Keane, the Science Use of NIF programme, C. Wild (Fraunhofer Institute for Applied Solid-State

Physics, Freiburg, Germany) for preparation of the diamond targets, D. Hicks for his analysis work, and M. Millot for reanalysing published diamond Hugoniot data. This work was performed under the auspices of the US Department of Energy by Lawrence Livermore National Laboratory under contract number DE-AC52-07NA27344, with additional support from the Department of Energy, the University of California, and the Miller Institute for Basic Research in Science.

Author Contributions R.F.S., J.H.E., D.G.B., P.M.C., J.R.P., A.E.L. and G.W.C. designed, executed and analysed the data from the ramp compression experiments. J.H.E., R.E.R., L.X.B., R.J., T.S.D., J.W. and G.W.C. performed the comparisons of experimental data to EOS models and theory. J.B., T.B. and A.V.H. were instrumental in procuring and metrologizing the diamond step samples.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.W.C. (collins7@llnl.gov).

METHODS

Ramp-compression design to terapascal pressures. The inner wall of a gold hohlraum (a gold cylinder used to convert laser light to X-rays) was illuminated with 176 beams of the NIF with a combined energy up to 0.76 MJ in a ~ 20 -ns temporally ramped pulse. This generates a near-blackbody distribution of thermal X-rays with a characteristic radiation temperature T_r , which increases with time to a peak of $T_r \approx 235$ eV. The hohlraum was filled with 0.1 atmosphere of neopentayne (C_5H_{12}) gas, which enabled the hohlraum cavity to stay open so that input laser power could be coupled effectively at late times. The C_5H_{12} gas was held within the hohlraum by 0.6- μm -thick polyimide windows covering the laser entrance holes. The X-ray ablation of diamond produces a uniform ramp-compression wave that transits the diamond sample. As the compression wave reaches the back of the sample, the surface accelerates into free space, and the free-surface velocity history u_{fs} for each step is recorded with a line-imaging velocity interferometer (VISAR) (Fig. 1). Our laser pulse shape is designed to launch an initial elastic shock into the diamond sample in advance of the ramp-compression wave. This shock feature—observed in the free-surface velocity record at $u_{fs} = 4.1 \text{ km s}^{-1}$ (Fig. 1) and corresponding to $P_{x, \text{limit}} = 0.11 \text{ TPa}$ —is interpreted as the dynamic strength (elastic limit) of diamond. The corresponding dynamic yield strength Y_0 is determined from $Y_0 = P_{x, \text{limit}}(1 - 2\nu)/(1 - \nu)$, with the Poisson's ratio, $\nu = 0.18$, derived from our sound-speed data (Fig. 2a) from $\left(\frac{C_{\text{longitudinal}}}{C_{\text{bulk}}}\right)^2 = 3\left(\frac{1-\nu}{1+\nu}\right)$. This yields $Y_0 = 0.085 \text{ TPa}$, which is less than observed in static experiments³¹ ($Y_0 = 0.13\text{--}0.15 \text{ TPa}$) but consistent with the values $0.069 \text{ TPa} < Y_0 < 0.096 \text{ TPa}$ reported for ramp compression of diamond with micrometre grain size³². The presence of an initial shock results in a loss of diamond strength¹⁸, with expected lower levels of compressive work heating over pure ramp compression³² and, therefore, a lower-temperature compression path. **Target design.** Our samples consist of a 50- μm -thick diamond plate used as an ablator, a 10- μm -thick Au layer preheat shield, and a diamond plate having four steps (Fig. 1, inset). The diamond was synthesized by chemical vapour deposition to yield a layered microstructure with an average grain size of 200 nm and a density of 3.2491 g cm^{-3} ($\pm 0.01\%$)^{33, 34}. The final sample had alternating 0.35- μm -thick layers of 20-nm grains and ~ 350 -nm grains. X-ray diffraction showed a $\langle 110 \rangle$ texture in the growth direction. The thickness of the sample is determined to $\pm 0.1 \mu\text{m}$, including uncertainties in the diamond ablator and Au thicknesses, whereas the differences in step thickness are determined by optical interferometry to $\pm 0.1 \mu\text{m}$. The diamond sample was then attached to the Au with a ~ 3 - μm -thick glue layer. The Au layer was incorporated into the target design to serve as a radiation preheat shield. Detailed radiation transport simulations estimate a temperature rise of 33 K, due to X-ray preheating.

Velocity interferometry. The response of the sample is characterized by velocity interferometry (VISAR), which records the velocity of the sample's free (back) surface as it is engulfed by the pressure wave (Fig. 1). The VISAR (Velocity Interferometer System for Any Reflector) diagnostic uses a line-focused 660-nm-wavelength laser beam to monitor a ~ 1 -mm strip across all four steps of the sample³⁵. Changes in velocity of the diamond free surface produce phase shifts in interference fringes that are recorded with a streak camera (Fig. 1). A typical VISAR record has a 30- μm spatial resolution, a 10-ns streak window with 0.01-ns resolution, and a velocity resolution of 0.1 km s^{-1} .

Stress–density analysis. Iterative Lagrangian analysis is used to translate these velocity data into a stress–density relation that quantifies the loading path (Fig. 2)^{17, 36}. The Lagrangian analysis method developed by Aidun and Gupta³⁶ and modified by Rothman¹⁷ was used to determine the Lagrangian sound speed $C_L(u)$ and the stress–density ($P_x - \rho$) relation from the measured $u_{fs}(t)$ data, where u is the particle speed, and u_{fs} is the sample's free surface velocity (across each of four thicknesses). Metrology of the sample surface showed that the roughness was $< 0.1 \mu\text{m}$, thickness gradients were $< 1\%$, and step heights were accurate to within $0.1 \mu\text{m}$. In all, three shots gave $C_L(u)$ and $P_x - \rho$ data. $C_L(u)$ and its uncertainty $\sigma_{C_L}(u)$ are obtained from thickness and velocity-versus-time data by linear regression using errors determined by our measurement accuracies: u_{fs} (0.05 km s^{-1}), time (10 ps), and step height (100 nm). The uncertainty is propagated by calculating the weighted mean average of all three shots, $C_L(u) = \sum_j \frac{C_{L,j}}{\sigma_{C_{L,j}}^2} / \sum_j \frac{1}{\sigma_{C_{L,j}}^2}$, as shown by the blue curve in Fig. 2a, where j is the shot number. The uncertainty in the average value is chosen from the maximum of the uncertainty in the mean and the weighted standard deviation. $C_L(u)$ and σ_{C_L} are integrated to obtain $P_x = \rho \int_0^u C_L du$, $\rho = \rho_0 \left(1 - \int_0^u \frac{du}{C_L}\right)^{-1}$, and their uncertainties $\sigma_{P_x} = \rho_0 \int_0^u \sigma_{C_L} du$ and $\sigma_\rho = \frac{\rho^2}{\rho_0} \int_0^u \frac{\sigma_{C_L}}{C_L^2} du$. Uncertainties are propagated through the integrals linearly, rather than in quadrature, because σ_{C_L} appears to be strongly correlated rather than random. This method of uncertainty

propagation allows the direct propagation of experimental uncertainties to $P_x - \rho$. Sound speed analysis over the three steps (four thicknesses) show simple wave behaviour, suggesting that the material response is not time-dependent within the experimental uncertainties.

Release waves from the diamond–vacuum interface significantly perturb the incoming ramp wave. Extensive tests using simulated data confirm that the iterative Lagrangian analysis accurately corrects for these wave interactions.

Mie–Grüneisen Hugoniot and cold curve. We compare our stress–density data (Fig. 2b and Extended Data Fig. 1) to a Hugoniot and cold curve reduced from available diamond Hugoniot data. There are several ways to construct a Mie–Grüneisen EOS, and here we begin with the relation for the pressure relative to a reference pressure P_{ref}

$$P(\eta, E) = P_{\text{ref}}(\eta) + \rho_0 \eta^\gamma (E - E_{\text{ref}}(\eta)) \quad (1)$$

where $\eta = \frac{\rho}{\rho_0}$ is the compression, γ is the Grüneisen parameter (assumed to depend only on density) and ρ_0 is the initial density. We can use either the Hugoniot or isotherm data to determine the reference states. Here we use the diamond Hugoniot data as the reference using a linear fit to existing shock velocity versus particle velocity data^{18, 37–40}

$$Us = C + sUp \quad (2)$$

where $C = 12.0 \text{ km s}^{-1}$ and $s = 1.04$. From this we obtain

$$P_{\text{ref}}(\eta) = P_H(\eta) = \rho_0 \eta \frac{C^2(\eta - 1)}{(\eta - s(\eta - 1))^2} \quad (3)$$

$$E_{\text{ref}}(\eta) = E_H(\eta) = \frac{C^2(\eta - 1)^2}{2(\eta - s(\eta - 1))^2} \quad (4)$$

where $P_H(\eta)$ and $E_H(\eta)$ are the Hugoniot pressure and energy, respectively. Finally, from equation (1) we obtain the cold curve

$$P_0(\eta) = \rho_0 \eta \left(\frac{C^2(\eta - 1)}{(\eta - s(\eta - 1))^2} + \gamma \left(E_0 - \frac{C^2(\eta - 1)^2}{2(\eta - s(\eta - 1))^2} \right) \right) \quad (5)$$

where we solve for $E_0(\eta)$ by

$$\begin{aligned} \frac{dE_0}{d\eta} &= \frac{1}{\rho_0 \eta^2} (P_H + \rho_0 \gamma \eta (E_0 - E_H)) \\ &= \frac{1}{\eta} \left(\frac{C^2(\eta - 1)}{(\eta - s(\eta - 1))^2} + \gamma \left(E_0 - \frac{C^2(\eta - 1)^2}{2(\eta - s(\eta - 1))^2} \right) \right) \end{aligned} \quad (6)$$

It is also assumed $\gamma = \gamma_0 \eta^{-q}$, where $\gamma_0 = 0.85$ (ref. 21). The variable q has not been measured at high pressure, and can have a significant impact on the cold curve determined. We find that a value of $q = 1$ yields a cold curve centred on the DFT-calculated cold curve¹². This value of q is consistent with static measurements at pressures $< 0.1 \text{ TPa}$ (ref. 21). This simple model for calculating the cold curve does not incorporate volume changes from proposed high-pressure phase transformations.

Calculation of 7.6% porous Hugoniot. The calculation is as shown in Fig. 2 and Extended Data Fig. 1. Our samples had a measured ambient density of 3.249 g cm^{-3} which is 7.6% below full crystal density. To calculate the stress–density path of a 7.6% porous Hugoniot we use the expression of McQueen⁴¹

$$P_x^*(\rho) = P_H \frac{1 - (\gamma/2) \left(\frac{\rho}{\rho_0} - 1 \right)}{1 - (\gamma/2) \left(\frac{\rho}{\rho_0^*} - 1 \right)} \quad (7)$$

where $P_x^*(\rho)$ is the stress state along the porous Hugoniot at a density ρ , ρ_0 is the initial full crystal density (3.515 g cm^{-3}), ρ_0^* is the initial porous density (3.249 g cm^{-3}) and $\gamma(\rho)$ is the Grüneisen parameter. We note that implicit within the porous Hugoniot expression in equation (7) is that the wave is steady and the pores have collapsed completely in the post-shock state, that is, $P_x^*(\rho) = 0$ for $\rho_0^* \rightarrow \rho_0$; an assumption which is incorrect for diamond. Equation (7) is therefore a poor estimate for weak shocks but in cases where the shock pressure greatly exceeds the material strength (after the pores have closed) it is reasonable.

Upon compression, the material strength determines how much stress is needed to reduce the porosity to a given level. This relationship can be summarized in a crush-up curve: $\rho = \rho(\rho_0^*, P_x^*, E)^{42-44}$. Following Carroll and Holt⁴³, pore crush-up is only initiated after a critical longitudinal stress, $P_{\text{crit}} = \frac{2}{3} Y_0 |\ln f_0|$ where Y_0 is the yield strength and f_0 is the initial porosity. For our diamond samples $Y_0 = 0.085 \text{ TPa}$, $f_0 = \rho_0^*/\rho_0 = 0.076$ and $P_{\text{crit}} = 0.146 \text{ TPa}$. For $0 \leq P \leq P_{\text{crit}}$, the pressure-dependent

pore fraction $f = f_0$ and the material is assumed to deform elastically. For $P > P_{\text{crit}}$, the porosity decays exponentially as $f = e^{-3P_c/2Y_0}$.

A number of studies on shock compression of under-dense materials have shown that rapid heating due to pore closure and the resultant increase in thermal pressure gives rise to reduced compression⁴¹. In Extended Data Fig. 1 this is witnessed by the stiffer response of the calculated porous Hugoniot compared to the Hugoniot for full-density diamond.

Diamond EOS data and DFT calculations. Extended Data Fig. 1 compares our data (initial density $\rho_0 \approx 3.249 \text{ g cm}^{-3}$) with previously reported shock Hugoniot^{18,37–40}, static²², and ramp compression³² data ($\rho_0 \approx 3.515 \text{ g cm}^{-3}$) as stress versus density. Shock Hugoniot data rely on knowledge of a reference material and therefore subsequent revisions of the reference EOS can change the reported diamond Hugoniot data. The Hugoniot points shown in Extended Data Fig. 1 have been reanalysed to account for new standard EOS as follows: The data reported by Nagao³⁸ and four of the high pressure points of Hicks⁴⁰ (open pentagons) used aluminium as a standard and were reanalysed using impedance matching⁴⁵ with the latest fit to the aluminium Hugoniot⁴⁶. The highest pressure point of Hicks used a Mo standard and remains unchanged. Additional data reported by Hicks⁴⁰ and data reported by Brygoo³⁹ used a quartz standard. These data have been reanalysed using the constant Grüneisen re-shock model in ref. 40 and the quartz Hugoniot used as a reference is a fit of all available data for quartz shocked into the liquid phase^{46,47}.

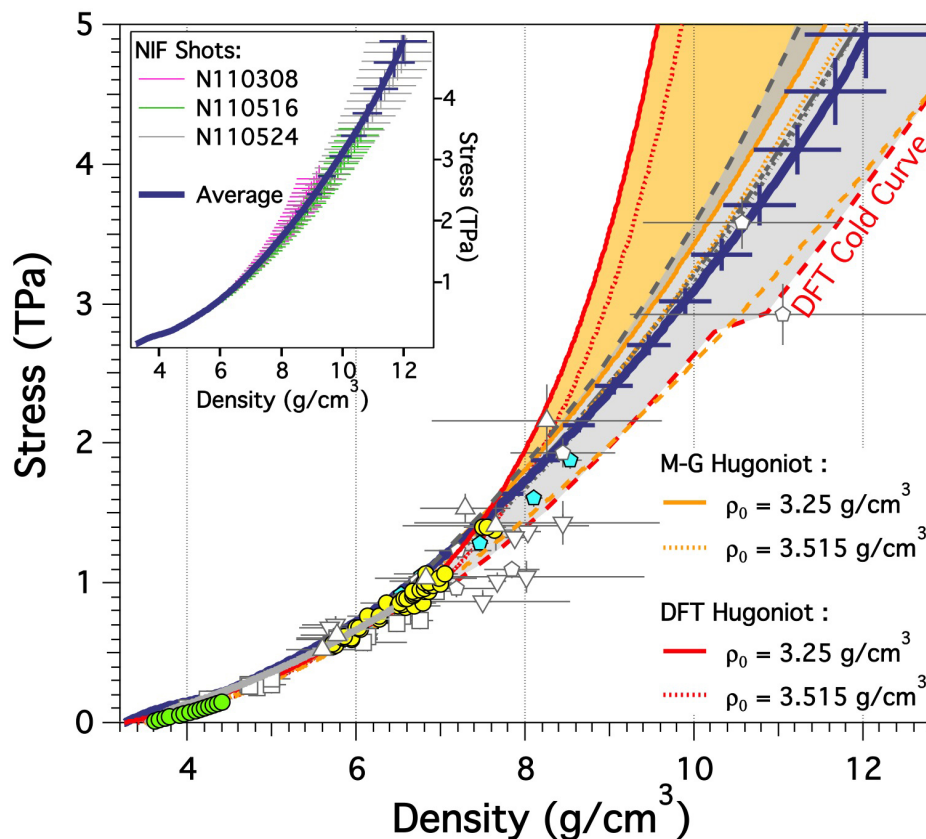
The DFT EOS we use to produce the Hugoniot in Fig. 2 and Extended Data Fig. 1 is as reported¹⁰, except without the embedding into the Thomas–Fermi-based quotidian-EOS (QEOS) model. We omit the connection with the QEOS model because the transition region between *ab initio* and QEOS models in ref. 10 created unphysical kinks in the EOS and resulting Hugoniot. The extrapolation of the more limited-range *ab initio* EOS of ref. 10 to the conditions relevant for the Hugoniot final states shown in our figures is expected to be quite accurate⁴⁸. The DFT cold curve generated from ref. 10 is in good agreement with the DFT cold curve reported in ref. 12 (red dashed curve in Fig. 2 and Extended Data Fig. 1) for stresses less than 2.5 TPa (which is the pressure below which *ab initio* electronic structure information was used to construct that EOS).

Static-compression and elasticity measurements to 0.15 TPa are indistinguishable from the cold curves presented here^{21,22}. The fit to the static compression measurements over this low compression range ($\rho/\rho_0 \approx 1.18$) are insensitive to the form of EOS used to fit the data (for example, Vinet¹⁹, Birch–Murnaghan²⁰, or Holzapfel⁴⁹). The Vinet EOS plotted in Fig. 2 and Extended Data Fig. 1 use $K_0 = 445 \text{ GPa}$ and $K'_0 = 4.18$ as reported in ref. 21. The values used for the Birch–Murnaghan ($K_0 = 445 \text{ GPa}$, $K'_0 = 3.90(0.04)$) and Holzapfel ($K_0 = 445 \text{ GPa}$, $K'_0 = 3.95(0.05)$) forms of EOS are based on fits to previous isothermal data^{21,22}. Here the values from ref. 22 have been reanalysed using the revised ruby pressure scale as reported in ref. 21. Extrapolating these isothermal data to the multi-terapascal regime becomes highly uncertain depending on the EOS used (Fig. 2b and Extended Data Fig. 1).

Although temperature was not measured in these experiments, it is useful to comment on such estimates from theoretical calculations. The temperature calculated

from DFT along the diamond principal isentrope is quite low even at the most extreme compressions studied here ($\sim 600 \text{ K}$ to 700 K at multi-terapascal pressures). For this reason, the principal isentrope and the room-temperature isotherm are predicted to be nearly coincident in stress–density space. It is certainly possible that our ramp compression path have higher temperatures than these isentrope predictions and this may be responsible for the higher stress versus density. However, because temperature, material strength¹⁸, and phase transformation kinetics²³ can each cause a stiffer response with respect to the isentrope, current estimates for the ramp compression temperature into the terapascal regime are quite speculative.

31. Eremets, M. I. *et al.* The strength of diamond. *Appl. Phys. Lett.* **87**, 141902 (2005).
32. Bradley, D. K. *et al.* Diamond at 800 GPa. *Phys. Rev. Lett.* **102**, 075503 (2009).
33. Reichart, P. *et al.* Three-dimensional hydrogen microscopy in diamond. *Science* **306**, 1537–1540 (2004).
34. Daweideit, C. *et al.* Grain size dependent physical and chemical properties of thick CVD diamond films for high energy density physics experiments. *Diamond Rel. Mater.* **40**, 75–81 (2013).
35. Celliers, P. M. *et al.* Line-imaging velocimeter for shock diagnostics at the Omega laser facility. *Rev. Sci. Instrum.* **75**, 4916–4929 (2004).
36. Aidun, J. B. & Gupta, Y. M. Analysis of Lagrangian gauge measurements of simple and nonsimple plane waves. *J. Appl. Phys.* **69**, 6998–7014 (1991).
37. Knudson, M. D., Desjarlais, M. P. & Dolan, D. H. Shock-wave exploration of the high-pressure phases of carbon. *Science* **322**, 1822–1825 (2008).
38. Nagao, H. *et al.* Hugoniot measurement of diamond under laser shock compression up to 2 TPa. *Phys. Plasmas* **13**, 052705 (2006).
39. Brygoo, S. *et al.* Laser-shock compression of diamond and evidence of a negative-slope melting curve. *Nature Mater.* **6**, 274–277 (2007).
40. Hicks, D. G. *et al.* High-precision measurements of the diamond Hugoniot in and above the melt region. *Phys. Rev. B* **78**, 174102 (2008).
41. McQueen, R. G., Marsh, S. P., Taylor, J. W., Fritz, J. N. & Carter, W. J. in *High Velocity Impact Phenomena* Ch. 7, 293–417 (Academic Press, 1970).
42. Herrmann, W. Constitutive equation for the dynamic compaction of ductile porous materials. *J. Appl. Phys.* **40**, 2490–2499 (1969).
43. Carroll, M. M. & Holt, A. C. Static and dynamic pore-collapse relations for ductile porous materials. *J. Appl. Phys.* **43**, 1626–1636 (1972).
44. Reisman, D. B., Wolfer, W. G., Elsholz, A. & Furnish, M. D. Isentropic compression of irradiated stainless steel on the Z accelerator. *J. Appl. Phys.* **93**, 8952–8957 (2003).
45. Celliers, P. M. *et al.* Systematic uncertainties in shock-wave impedance-match analysis and the high-pressure equation of state of Al. *J. Appl. Phys.* **98**, 113529 (2005).
46. Knudson, M. D. & Desjarlais, M. P. Adiabatic release measurements in α -quartz between 300 and 1200 GPa: characterization of α -quartz as a shock standard in the multimegabar regime. *Phys. Rev. B* **88**, 184107 (2013).
47. Hicks, D. G. *et al.* Shock compression of quartz in the high-pressure fluid regime. *Phys. Plasmas* **12**, 082702 (2005).
48. Benedict, L. X. *et al.* A multiphase equation of state for carbon addressing high pressures and temperatures. Preprint at <http://arxiv.org/abs/1311.4577> (2013).
49. Holzapfel, W. Equations of state for ideal and real solids under strong compression. *Europhys. Lett.* **16**, 67 (1991).



Extended Data Figure 1 | Ramp-compressed diamond stress versus density compared to other high-pressure data. NIF ramp-compression data with 1σ error bars (solid blue line) together with calculated Hugoniot (low-initial-density diamond, solid red line; standard-initial-density diamond, dotted red line) and the calculated cold curve (dashed red line)¹² from DFT; a simple Mie–Grüneisen model reduction of Hugoniot data to produce an extrapolated Hugoniot (low-initial-density diamond, solid orange line; standard-initial-density diamond, dotted orange line), and cold curve (dashed orange line); Vinet¹⁹ (dot-dashed grey line), Birch–Murnaghan²⁰ (dashed grey line), and Holzapfel⁴⁹ (dotted grey line) extrapolations of 300-K diamond anvil cell data^{21,22}. The shaded regions show the range of different models for cold curve

(grey) and Hugoniot (orange) showing roughly the range of uncertainty in this ultrahigh-pressure regime. Also shown are data from shock experiments (yellow circles³⁷, up triangles³⁸, open pentagons (which used an Al or Mo standard)⁴⁰, down triangles³⁹, blue pentagons (which used the more accurate quartz standard)⁴⁰, open squares¹⁸), isothermal static data (green circles are ruby-corrected data^{21,22}) and the ramp-compression data of Bradley³² (solid grey line). The ramp-compression data of Bradley used full-density diamond and did not use an initial shock as in NIF data. The inset shows the calculated stress–density relations of the three NIF shots: N110308, N110516 and N110524, showing the level of repeatability between experiments.

Extended Data Table 1 | Ramp-compressed diamond stress–density data

P_x (GPa)	σ_{P_x} (GPa)	ρ (g/cm ³)	σ_ρ (g/cm ³)	P_x (GPa)	σ_{P_x} (GPa)	ρ (g/cm ³)	σ_ρ (g/cm ³)
0	0	3.25	0	1422	25	7.47	0.13
39.2	0.3	3.4	0	1487	27	7.58	0.14
79	0.7	3.57	0	1554	29	7.7	0.14
98.8	0.8	3.66	0	1623	31	7.82	0.15
117	1	3.76	0.01	1694	33	7.94	0.16
133.8	1.1	3.88	0.01	1766	35	8.06	0.17
151	1.3	4	0.01	1839	38	8.18	0.18
168.7	1.4	4.13	0.01	1915	40	8.31	0.19
187.8	1.6	4.25	0.01	1992	43	8.43	0.2
210	1.9	4.37	0.01	2071	46	8.55	0.21
234	2.1	4.48	0.02	2152	49	8.68	0.22
261	2.4	4.58	0.02	2235	52	8.8	0.23
289	2.8	4.69	0.02	2320	55	8.93	0.24
318	3.2	4.79	0.02	2406	59	9.05	0.25
348	3.6	4.9	0.02	2495	62	9.18	0.27
379	4	5	0.03	2585	66	9.31	0.28
412	4.5	5.11	0.03	2678	70	9.44	0.29
445	5	5.22	0.03	2774	74	9.57	0.3
480	5.5	5.33	0.04	2871	79	9.7	0.32
516	6	5.44	0.04	2969	85	9.83	0.34
553	6.6	5.55	0.04	3069	92	9.96	0.36
592	7.3	5.66	0.05	3170	100	10.1	0.38
632	7.9	5.77	0.05	3273	108	10.23	0.4
674	8.7	5.88	0.05	3379	116	10.37	0.42
717	9.5	5.99	0.06	3486	125	10.51	0.44
762	10	6.1	0.06	3596	134	10.65	0.46
808	11	6.21	0.07	3710	144	10.78	0.49
856	12	6.32	0.07	3827	155	10.92	0.51
905	13	6.44	0.07	3947	165	11.06	0.54
955	14	6.55	0.08	4070	177	11.19	0.56
1008	15	6.66	0.09	4196	189	11.33	0.59
1062	17	6.78	0.09	4324	202	11.47	0.62
1118	18	6.89	0.1	4457	215	11.6	0.65
1176	19	7	0.1	4592	242	11.74	0.69
1235	21	7.12	0.11	4736	275	11.87	0.74
1296	22	7.23	0.12	4889	313	12	0.79
1358	24	7.35	0.12	4929	323	12.03	0.8

Tabulated data showing stress (P_x), stress uncertainty (σ_{P_x}), density (ρ) and density uncertainty (σ_ρ). All uncertainties are 1σ .

A low-cost non-toxic post-growth activation step for CdTe solar cells

J. D. Major¹, R. E. Treharne¹, L. J. Phillips¹ & K. Durose¹

Cadmium telluride, CdTe, is now firmly established as the basis for the market-leading thin-film solar-cell technology. With laboratory efficiencies approaching 20 per cent¹, the research and development targets for CdTe are to reduce the cost of power generation further to less than half a US dollar per watt (ref. 2) and to minimize the environmental impact. A central part of the manufacturing process involves doping the polycrystalline thin-film CdTe with CdCl₂. This acts to form the photovoltaic junction at the CdTe/CdS interface^{3,4} and to passivate the grain boundaries⁵, making it essential in achieving high device efficiencies. However, although such doping has been almost ubiquitous since the development of this processing route over 25 years ago⁶, CdCl₂ has two severe disadvantages; it is both expensive (about 30 cents per gram) and a water-soluble source of toxic cadmium ions, presenting a risk to both operators and the environment during manufacture. Here we demonstrate that solar cells prepared using MgCl₂, which is non-toxic and costs less than a cent per gram, have efficiencies (around 13%) identical to those of a CdCl₂-processed control group. They have similar hole densities in the active layer ($9 \times 10^{14} \text{ cm}^{-3}$) and comparable impurity profiles for

Cl and O, these elements being important p-type dopants for CdTe thin films. Contrary to expectation, CdCl₂-processed and MgCl₂-processed solar cells contain similar concentrations of Mg; this is because of Mg out-diffusion from the soda-lime glass substrates and is not disadvantageous to device performance. However, treatment with other low-cost chlorides such as NaCl, KCl and MnCl₂ leads to the introduction of electrically active impurities that do compromise device performance. Our results demonstrate that CdCl₂ may simply be replaced directly with MgCl₂ in the existing fabrication process, thus both minimizing the environmental risk and reducing the cost of CdTe solar-cell production.

The cost of CdTe photovoltaic modules has now dropped below one US dollar per watt and the cost of power generation is rapidly approaching grid parity². A key stage in the fabrication of CdTe solar cells is to anneal the CdTe/CdS p-n structure in the presence of CdCl₂. Widely referred to as the 'activation' step, this converts a cell with <2% conversion efficiency to one with typically >10% efficiency and is linked to a number of beneficial structural and electrical changes in both the CdTe and CdS layers³⁻⁶. CdCl₂ contributes to electrical doping, the

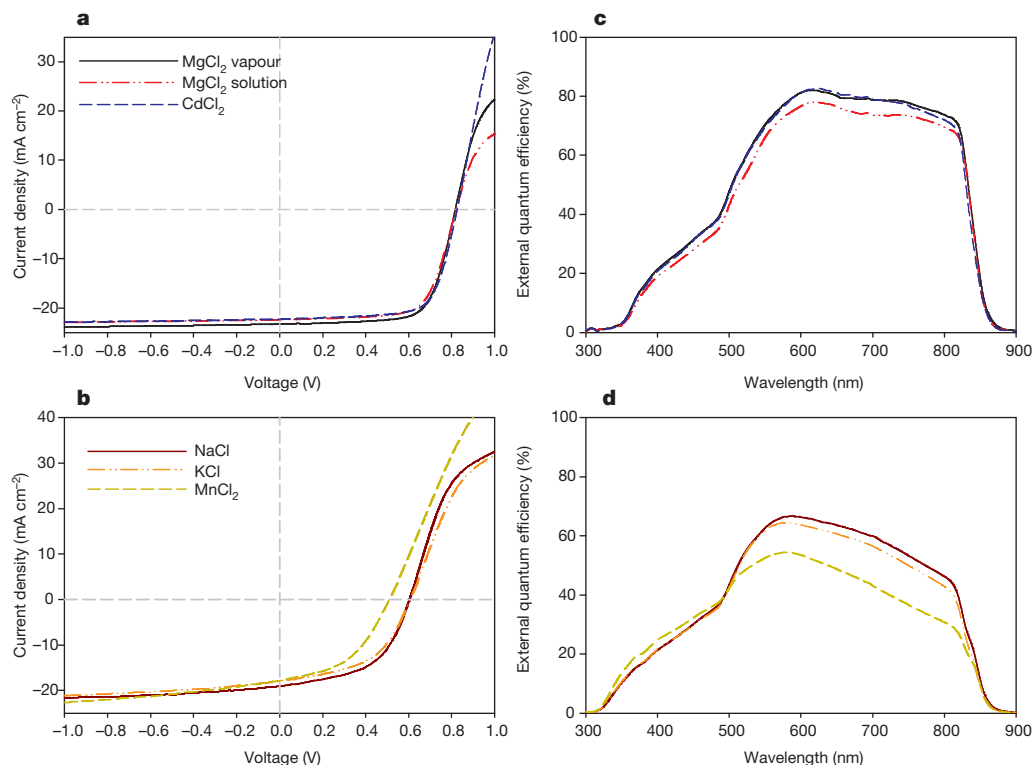


Figure 1 | *J*-*V* and EQE analysis of cells with different chloride treatments. *J*-*V* curves for the highest-efficiency contacts for MgCl₂-vapour-treated, MgCl₂-solution-treated and CdCl₂-treated devices (a) and for cells activated using the alternative low-cost chlorides, NaCl, KCl and MnCl₂ (b). All curves

show a high degree of 'roll-over' in forward bias. EQE curves for the highest-efficiency contacts for MgCl₂-vapour-treated, MgCl₂-solution-treated and CdCl₂-treated devices (c) and for NaCl, KCl and MnCl₂ cells that show a decrease in EQE values from short to long wavelength (d).

¹Stephenson Institute for Renewable Energy and Department of Physics, School of Physical Sciences, Chadwick Building, University of Liverpool, Liverpool L69 7ZF, UK.

Table 1 | Peak solar-cell performance for all chlorides tested

Treatment	Peak efficiency (%)	Peak fill factor (%)	Peak J_{sc} (mA cm^{-2})	Peak V_{oc} (V)
CdCl ₂	13.02	70.01	22.13	0.831
MgCl ₂ solution	12.71	69.08	22.41	0.821
MgCl ₂ vapour	13.50	70.24	23.26	0.826
NaCl	6.75	53.34	19.78	0.603
MnCl ₂	4.37	45.87	18.30	0.520
KCl	5.49	50.11	17.95	0.607

Key solar-cell performance parameters—efficiency η , fill factor, short-circuit current density J_{sc} and open-circuit voltage V_{oc} —for the highest-efficiency contacts for the various activation treatments tested.

recrystallization of small grains and to the passivation of grain boundaries and interface states. The two key drivers in CdTe device research are to limit the environmental impact and to reduce the cost of production. The use of CdCl₂ is problematic for both. While CdTe and CdS are both stable, insoluble and are anticipated to contribute little Cd to the environment⁷, CdCl₂ powder is highly toxic and water soluble, thus posing a risk to both industrial operators and the environment.

CdCl₂ also represents a large but inherently avoidable production cost. The bulk cost of CdCl₂ is about 0.3 US\$ per gram and it requires about 5 tonnes of CdCl₂ per gigawatt of solar-cell production, giving an estimated total cost of US\$1,500,000 per gigawatt. However, the largest cost associated with CdCl₂ processing lies in its handling and disposal, which require a specialized industrial plant for the protection of operators and specialist waste disposal.

However, despite its disadvantages, the use of CdCl₂ has endured for more than 25 years⁶ and comparatively little effort has been made to identify an effective replacement. A notable exception was the use of the chlorofluorocarbon gas HCF₂Cl (difluorochloromethane)⁸, which yielded high-efficiency devices. However, this too posed problems because the gas is linked to ozone depletion and its use has since been restricted by international agreements. Because a viable alternative has never been identified, CdCl₂ remains universal in commercial high-efficiency CdTe device production.

Here, we demonstrate that MgCl₂ may be used as a direct replacement for CdCl₂ in CdTe device manufacturing with no loss in cell performance. MgCl₂ has <1% of the cost per weight (about 0.001 US\$ per gram) of CdCl₂ and is recoverable from sea water⁹. It is also non-hazardous, environmentally safe and is already used widely—for example, in cold-weather road treatment, as a bath salt and as a food additive in the production of tofu¹⁰. A process change from CdCl₂ to MgCl₂ has huge potential instantly to reduce the cost of power generation by CdTe photovoltaics and to minimize the risks in industrial production.

In these experiments, a number of low-cost chlorides (MgCl₂, NaCl, KCl and MnCl₂) were compared in like-for-like CdTe solar-cell fabrication and performance tests. Other chlorides, which represented either an environmental risk or a high cost (such as CuCl₂ and ZnCl₂) were not considered. MgCl₂ produced the highest efficiencies and was therefore compared more extensively to a standard CdCl₂ control process. MgCl₂ was applied in two different variations of the basic process, in which the surface of the CdTe is first exposed to the chloride, and then annealed in a tube furnace—see Methods for further details. (1) In the ‘solution’ process MgCl₂ was applied directly to the free CdTe surface in saturated solution in methanol. The samples were then dried to form a layer, and annealed. (2) In the ‘vapour’ process a glass slide coated with MgCl₂ was placed alongside the solar-cell samples directly in the annealing furnace (vapour transport to the CdTe surface occurred during the annealing step itself). Apart from the chloride treatment step, all other device processing was identical.

Current density versus voltage (J – V) curves from the highest-efficiency devices measured under a simulated AM1.5 spectrum are shown in Fig. 1a for the MgCl₂ and CdCl₂ treatments, with those for other chlorides shown in Fig. 1b. Their external quantum efficiency (EQE) curves are given in Fig. 1c and d, respectively.

The most efficient single device measured (Table 1) was for the MgCl₂ vapour treatment. It had an efficiency of 13.50%, a fill factor of 70.24%, a short-circuit current density J_{sc} of 23.36 mA cm^{-2} and an open circuit voltage V_{oc} of 826 mV. The highest efficiency for any CdCl₂ control device was 13.02%. However, average performances measured over nine cells (Table 2) showed the CdCl₂ and MgCl₂-vapour treatments to give identical results within the margin of error, and the MgCl₂ solution treatment to yield only slightly reduced efficiencies.

For treatments using NaCl, KCl and MnCl₂, the best solar energy conversion efficiencies were all <6.7% owing to the low open circuit voltages and fill factor values that were associated with pronounced forward bias current limitation or ‘roll-over’. These efficiencies were less than half of that of the CdCl₂ control and MgCl₂ treatments.

It is often the case that at high forward bias the current in J – V curves for CdTe cells is depressed by the presence of a non-ohmic contact; this is usually referred to as ‘roll-over’¹¹. This occurs due to the very high electron affinity of CdTe, meaning that a metal contact to p-type CdTe always forms a Schottky barrier. This ‘roll-over’ can be seen to some extent for all samples measured, but was extremely pronounced in the NaCl-, KCl- and MnCl₂-treated samples. Some additional ‘roll-over’ is also visible for MgCl₂ treatment in comparison to CdCl₂ treatment. It has been suggested that the inclusion of a Cd_{1-x}Mg_xTe layer at the CdTe surface may improve the back-contact ohmicity¹². However, in the present samples, our analyses did not indicate the presence of any Cd_{1-x}Mg_xTe formation after MgCl₂ treatment and ‘roll-over’ was indeed present. In this case we attribute its presence to the formation of oxychloride surface phases similar to those which have been observed following CdCl₂ treatment¹³ and which increase contact resistance. Measurement of the back-contact barrier height, through J – V as a function of temperature¹⁴ (J – V – T), shows that the barrier height ϕ_b is slightly increased from 0.28 eV for CdCl₂ to 0.32 eV for MgCl₂ vapour treatment (Extended Data Fig. 1). However, through addition of a 2-nm-thick Cu layer to the back contact the barrier can be reduced to 0.23 eV, which is not anticipated to hinder device performance greatly^{11,15}.

The longer-term stability of MgCl₂-treated devices compared to CdCl₂-treated devices was also compared by J – V measurement both immediately after deposition and after a 6-month interval (Extended Data Fig. 2). Both devices were found to degrade identically (losing about 5% of their relative initial efficiency), consistent with the known degradation of the gold contacts used in laboratory-scale devices (that is, by oxidation of the underlying CdTe). No additional performance degradation related to MgCl₂ treatment was observed.

The EQE curve shapes show very little difference between the CdCl₂ and MgCl₂ treatments (Fig. 1b), indicating that the devices operate

Table 2 | Average solar-cell performance for CdCl₂- and MgCl₂-treated solar cells

Treatment	Average efficiency (%)	Average fill factor (%)	Average J_{sc} (mA cm^{-2})	Average V_{oc} (V)
CdCl ₂	12.97 ± 0.06	70.39 ± 0.78	22.14 ± 0.16	0.827 ± 0.009
MgCl ₂ solution	12.23 ± 0.42	69.10 ± 0.04	21.58 ± 0.73	0.820 ± 0.001
MgCl ₂ vapour	13.03 ± 0.67	71.20 ± 1.35	22.50 ± 1.07	0.818 ± 0.011

Average solar-cell performance parameters for MgCl₂ vapour-treated, MgCl₂ solution-treated and CdCl₂-treated devices. Results for the MgCl₂ vapour and CdCl₂ treatments agree within bounds of error. Average and standard deviation values are from batches of nine identical solar cells.

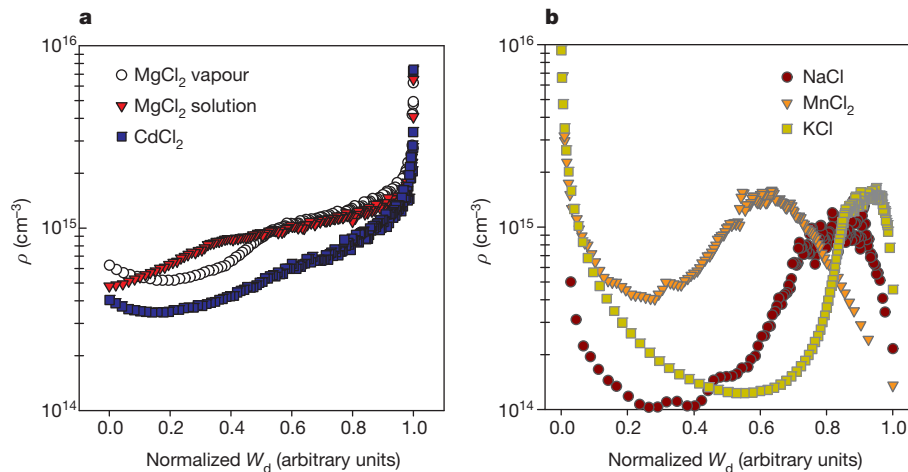


Figure 2 | Capacitance voltage profiling of carrier concentration. Hole density ρ versus normalized depletion width $W_d = \varepsilon\varepsilon_0 A/C$, determined from C - V measurements. The CdTe/CdS interface corresponds to $W_d = 0$ and $W_d = 1$ corresponds to the near-back surface point at which back-contact capacitance dominates ($V_{\text{bias}} = 500$ mV). **a**, Curves for optimized MgCl₂ and

identically and that there are no major differences in the junction position or recombination behaviour. On the other hand, the ineffective chlorides NaCl, KCl and MnCl₂ showed different behaviour (Fig. 1d), in which the EQE decreases from short to long wavelengths. This indicates either a decreased carrier lifetime or an increase in uncompensated impurities in these devices¹⁶ compared to CdCl₂ and MgCl₂ treatments.

Carrier density–depth profiles obtained from capacitance–voltage (C - V) data¹⁷ are shown in Fig. 2 for all devices. Here, the apparent hole density ρ is plotted as a function of the normalized depletion width W_d , calculated from the p - n junction capacitance $W_d = \varepsilon\varepsilon_0 A/C$, with A being the contact area, C the measured capacitance, ε the relative CdTe permittivity and ε_0 the permittivity of free space. $W_d = 0$ represents the position of the CdTe/CdS interface and $W_d = 1$ is the point at which the back-contact capacitance begins to dominate ($V_{\text{bias}} > 500$ mV).

It was found that the carrier density measured in the bulk of the films that was achieved with MgCl₂ is comparable to or greater than the bulk carrier density in films made using CdCl₂, the former giving $9 \times 10^{14} \text{ cm}^{-3}$ and the latter $5 \times 10^{14} \text{ cm}^{-3}$. Moreover, both MgCl₂-treated and CdCl₂-treated cells (Fig. 2a) show an increase in doping density towards the right-hand side of the plot, indicative of higher p doping at the back

CdCl₂ treatments. The carrier density is consistently high throughout the CdTe layer, increasing towards the back contact. **b**, Curves for ineffective treatments (NaCl, MnCl₂ and KCl), for which the carrier concentration has a peak in the CdTe bulk, low doping at the back.

surface¹⁷. This increase is beneficial to device performance because it will act to reduce the extent of the Schottky behaviour. A minor feature common to both treatments is a slight increase in carrier concentration near to the front contact that may be attributed to deep levels¹⁸.

For the ineffective treatments (Fig. 2b), the overall carrier concentrations are lower (less than $5 \times 10^{14} \text{ cm}^{-3}$), and their profile shapes are noticeably different. They show peaks in carrier concentration in the bulk of the film, and a reduction towards the back contact that will act to increase its Schottky barrier width and hence contribute to the performance loss from ‘roll-over’. The sharp increase in apparent doping at the near CdTe/CdS interface is an artefact of this Schottky contact: under high forward bias the contact junction undergoes a collapse, causing a decrease in capacitance that the plot interprets as a specious increase in carrier concentration¹⁹.

Secondary ion mass spectrometry (SIMS) analysis, shown in Fig. 3, was performed on MgCl₂ and CdCl₂ samples to measure the in-diffusion of Cl and O from the post-growth treatments, this being relevant since both are linked to p -type doping in CdTe^{17,20}. Indeed, the distribution of Cl and O in the depth profiles are qualitatively similar in shape to the carrier concentration profiles shown in Fig. 2 for both the MgCl₂-treated

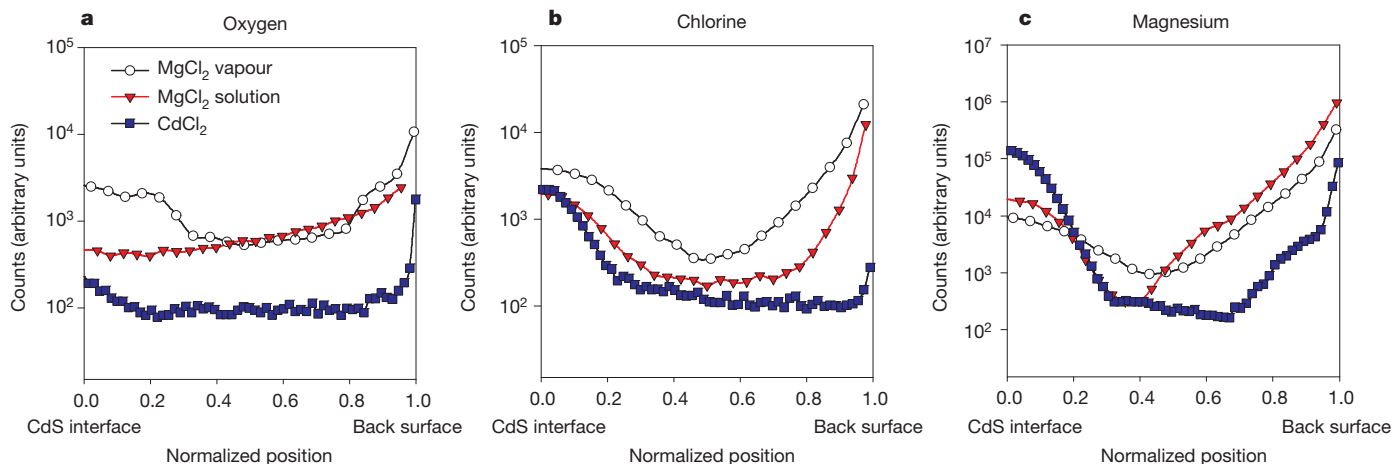


Figure 3 | SIMS profiles of CdTe films. SIMS profiles for oxygen (**a**), chlorine (**b**) and magnesium (**c**) content in CdCl₂-treated, MgCl₂-solution-treated and MgCl₂-vapour-treated CdTe layers. Plots have been normalized for position, so that the CdTe/CdS interface corresponds to 0 while the device back surface corresponds to 1. Analysis shows an increase in both chlorine and oxygen

content in CdTe for MgCl₂ treatment compared to CdCl₂ treatment. The Mg depth profiles show that both the CdCl₂- and the MgCl₂-treated samples contain similar levels of Mg. We attribute this to out-diffusion of Mg from the soda-lime glass substrate, which contains MgO as an ingredient.

and CdCl₂-treated samples. However, it is also noticeable that the MgCl₂ treatments result in a large increase in both the Cl and O content of the CdTe layers compared to CdCl₂ treatment. It is likely that this is the cause of the higher p-type doping measured for MgCl₂ treatments.

The SIMS depth profiles for Mg in Fig. 3 show the surprising result that the levels for both the CdCl₂ (having no intentionally introduced Mg) and the MgCl₂-treated devices were comparable. This arises due to Mg out-diffusion from the soda lime glass substrate, which is known to contain about 4% MgO. Contrary to expectation, MgCl₂ therefore introduces no new foreign impurities to the solar cells, when compared to existing practice. It is likely that MgCl₂ and CdCl₂ have comparable doping performance on account of the electrical similarity of their doubly charged cations. On the other hand, both the singly charged ions of Na and K, and the multiple oxidation states of Mn may be expected to be electrically active centres. Their use yields altered CdTe doping profiles (Fig. 3b) and this limits the device performance.

Further to the comparative cell data provided here, improved cell performance for MgCl₂ activation has been achieved through a series of cell and process developments. The ZnO buffer layer was replaced with a CdS:O nanostructured film²¹ and the CdS layer thickness reduced to about 40 nm. A 1 M MgCl₂ solution in water deposited via spray coating was used for treatment and a 2-nm Cu layer was added to improve the device back contact. This yielded a device of 15.7% efficiency (Extended Data Fig. 3), notably with a V_{oc} of 0.857 V, equivalent to that of the current CdTe champion cell¹. This clearly demonstrates that MgCl₂ is capable of producing high-efficiency devices.

These results demonstrate that MgCl₂ may be used as a direct replacement for CdCl₂ in the established activation process and as such is capable of instantly reducing the cost of CdTe solar-cell production. It also eliminates the need to use any water-soluble cadmium salt. It gives carrier concentrations and doping profiles that are desirable for high-efficiency devices and its effectiveness stems from the Mg²⁺ cation being electrically inactive in CdTe, unlike other low-cost chlorides investigated. A number of factors, such as longer-term module stability testing²², will still need to be assessed before industrial implementation. However, MgCl₂ processing has immense promise and has proved effective regardless of the manner in which it is applied. It is therefore likely to be robust to use under the conditions applicable for industrial application.

METHODS SUMMARY

Cell deposition. Cells were deposited on 'TEC10' soda lime glass coated with SnO₂:F supplied by NSG. 100-nm-thick ZnO buffer layers and a 120-nm-thick CdS layer were deposited by sputtering at room temperature and 200 °C respectively. CdTe layers were deposited by close space sublimation under 25 torr ambient nitrogen with source and substrate temperatures of 605 °C and 520 °C respectively. Prior to chloride treatment, cells were etched for 30 s in a nitric-phosphoric acid etch solution, followed by rinsing in deionized water. A second 30-s nitric-phosphoric etching step was applied following chloride treatment. Arrays of nine evaporated gold back-contacts (0.25 cm²) were then applied.

Treatment methods. Three main chloride process variations were compared: (1) the 'standard' CdCl₂ treatment, in which a 100-nm layer was evaporated onto the CdTe surface; (2) deposition of a 10% MgCl₂:90% methanol solution applied to the CdTe back surface; and (3) a 10% MgCl₂:90% methanol solution applied to a glass slide placed in the tube furnace alongside the sample. Tests for other chloride treatments using NaCl, KCl and MnCl₂ were performed via a 10% chloride:90% methanol solution applied to the CdTe surface. Annealing of samples was conducted in a tube furnace in air, with temperature and times being optimized in the range 390–450 °C and 10–60 min. Optimal treatment times were 20 min at 430 °C for MgCl₂ solution and CdCl₂ treatments, and 40 min at 430 °C for MgCl₂ vapour treatment.

Characterization. *J*–*V* analysis was performed under an AM1.5 spectrum at 1,000 W m^{−2} using a TS Space Systems solar simulator. *C*–*V* measurements were performed in the dark using a Solartron SI1260 impedance analyser. EQE measurements

were made using a Bentham PVE300 system. SIMS analysis was carried out by Loughborough Surface Analysis using a Cameca ims 4f instrument.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 February; accepted 28 April 2014.

Published online 25 June 2014.

- Green, M. A., Emery, K., Hishikawa, Y., Warta, W. & Dunlop, E. D. Solar cell efficiency tables (version 43). *Prog. Photovolt. Res. Appl.* **22**, 1–9 (2014).
- Zweibel, K. The impact of tellurium supply on cadmium telluride photovoltaics. *Science* **328**, 699–701 (2010).
- McCandless, B. E., Moulton, L. V. & Birkmire, R. W. Recrystallization and sulfur diffusion in CdCl₂-treated CdTe/CdS thin films. *Prog. Photovolt. Res. Appl.* **5**, 249–260 (1997).
- Moutinho, H. R. *et al.* Studies of recrystallization of CdTe thin films after CdCl₂ treatment. In *Conference Record of the Twenty Sixth IEEE Photovoltaic Specialists Conference (Anaheim, California)* 431–434 (IEEE, 1997).
- Edwards, P. R., Galloway, S. A. & Durose, K. EBIC and luminescence mapping of CdTe/CdS solar cells. *Thin Solid Films* **372**, 284–291 (2000).
- Basol, B. M., Ou, S. S. & Stafsudd, O. M. Type conversion, contacts, and surface effects in electroplated CdTe-films. *J. Appl. Phys.* **58**, 3809–3813 (1985).
- Fthenakis, V. M. & Kim, H. C. CdTe photovoltaics: life cycle environmental profile and comparisons. *Thin Solid Films* **515**, 5961–5963 (2007).
- Mazzamuto, S. *et al.* A study of the CdTe treatment with a freon gas such as CHF₂Cl. *Thin Solid Films* **516**, 7079–7083 (2008).
- Balarew, C. Solubilities in seawater-type systems—some technical and environmental friendly applications. *Pure Appl. Chem.* **65**, 213–218 (1993).
- Liu, Z. S. & Chang, S. K. C. Optimal coagulant concentration, soymilk and tofu quality as affected by a short-term model storage of proto soybeans. *J. Food Process. Preserv.* **32**, 39–59 (2008).
- Demtsu, S. H. & Sites, J. R. Effect of back-contact barrier on thin-film CdTe solar cells. *Thin Solid Films* **510**, 320–324 (2006).
- Drayton, J., Geisthardt, R., Raguse, J. & Sites, J. R. Metal chloride passivation treatments for CdTe solar cells. *Mater. Res. Soc. Symp. Proc.* **1538**, 269–274 (2013).
- Waters, D. *et al.* Surface analysis of CdTe after various pre-contact treatments. In *Proceedings of the 2nd World Conference and Exhibition of Photovoltaic Solar Energy Conversion* 1031–1034 (European Commission Joint Research Centre, 1995).
- Bätzner, D. L., Oszan, M. E., Bonnet, D. & Bucher, K. Device analysis methods for physical cell parameters of CdTe/Cds solar cells. *Thin Solid Films* **361/362**, 288–292 (2000).
- Fahrenbruch, A. L. Exploring back contact technology to increase CdS/CdTe solar cell efficiency. *Mater. Res. Soc. Symp. Proc.* **1012**, abstr. 1012-Y07-05 (2007).
- Kosyachenko, L. & Toyama, T. Current-voltage characteristics and quantum efficiency spectra of efficient thin-film CdS/CdTe solar cells. *Sol. Energy Mater. Sol. Cells* **120**, 512–520 (2014).
- Proskuryakov, Y. Y. *et al.* Doping levels, trap density of states and the performance of co-doped CdTe(As,Cl) photovoltaic devices. *Sol. Energy Mater. Sol. Cells* **93**, 1572–1581 (2009).
- Burgelman, M., Nollet, P. & Degraeve, S. Electronic behaviour of thin-film CdTe solar cells. *Appl. Phys. A* **69**, 149–153 (1999).
- Li, J. V. *et al.* Theoretical analysis of effects of deep level, back contact, and absorber thickness on capacitance-voltage profiling of CdTe thin-film solar cells. *Sol. Energy Mater. Sol. Cells* **100**, 126–131 (2012).
- Tyan, Y. S., Vazan, F. & Barge, S. Effect of oxygen on thin-film CdS/CdTe solar cells. In *Conference Record of the Twenty Sixth IEEE Photovoltaic Specialists Conference (Kissimmee, Florida)* 840–845 (IEEE, 1984).
- Wu, X. *et al.* Nanostructured CdS: O film: preparation, properties, and application. In *11th International Conference on II–VI Compounds Proceedings* 1062–1066 (Wiley-VCH, 2004).
- Barth, K. L., Enzenroth, R. A. & Sampath, W. S. Consistent processing and long term stability of CdTe devices. In *Conference Record of the Twenty First IEEE Photovoltaic Specialists Conference (Lake Buena Vista, Florida)* 323–326 (IEEE, 2005).

Acknowledgements We thank T. Veal for assistance in manuscript preparation and the Engineering and Physical Sciences Research Council for funding support.

Author Contributions J.D.M. and R.E.T. conceived the experiments. J.D.M. fabricated and tested the solar-cell devices. L.J.P. performed *C*–*V* measurements. J.D.M. and K.D. discussed the results and prepared the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.D.M. (jon.major@liverpool.ac.uk).

METHODS

Cell deposition. CdTe solar-cell devices were fabricated in the 'superstrate' configuration on commercial soda lime glass substrates (NSG TEC10 from NSG) coated with fluorine-doped tin oxide. Radio-frequency sputtering was used to deposit a 100-nm ZnO 'buffer' layer onto the fluorine-doped tin oxide by reactively sputtering from a Zn target in the presence of oxygen. A 120-nm-thick CdS layer window layer was then deposited by radio-frequency sputtering at a substrate temperature of 200 °C under ambient Ar and using a power of 60 W. CdTe absorber layers were deposited via close space sublimation deposition, using a custom all-quartz deposition chamber manufactured by Electro-Gas Systems. Deposition was carried out at source and substrate temperatures of 615 °C and 520 °C respectively, under a pressure of 30 torr of nitrogen, giving a thickness of about 4 µm. Following CdTe deposition, the samples were submerged for 15 s in a nitric-phosphoric acid etch solution. This created a Te-rich back surface to the CdTe layer and has been found to aid in-diffusion during post-growth activation²³. Following activation treatment (described below), samples were subjected to a further 15 s nitric-phosphoric etch before the deposition of 0.25 cm² gold back contacts by vacuum evaporation to complete the device.

Post growth activation treatments. A number of different post-growth treatment routes were compared in this work. Following deposition of the Cl-containing layer, all samples were annealed in a tube furnace under an air ambient. For each Cl treatment the annealing time and temperature was optimized in the range 10–60 min and 390–450 °C. This was done to ensure the maximum attainable performance level achievable by each treatment was accurately established. For each treatment time and temperature a complete device was fabricated and its performance was assessed via *J–V* analysis. The treatment which yielded the highest device efficiencies in each case was identified. Over 150 devices were processed during the course of this optimization.

For the standard CdCl₂ treatment, CdCl₂ (99.99% purity, Sigma Aldrich product number 202908) was deposited onto the CdTe back surface as a 100-nm-thick thin film via thermal evaporation, as this is the established deposition practice for cell production. Optimum treatment conditions were a 20-min anneal at 430 °C. Treatment using alternative chlorides—NaCl, KCl and MnCl₂—was performed via a few drops of 10% chloride: 90% methanol solution applied to the CdTe back surface before annealing. In the case of MgCl₂ (99+% purity, Alfa Aesar product number 12315), two different processing routes were initially investigated: (1) the MgCl₂ 'solution' treatment, in which a few drops of 10% MgCl₂:90% methanol solution were applied directly to the CdTe back surface; and (2) the MgCl₂ 'vapour' treatment, in which a few drops of 10% MgCl₂:90% methanol solution were applied to a glass slide placed in the tube furnace alongside the CdTe sample. Optimal processing conditions were found to be 20 min at 430 °C for MgCl₂ solution processing and 40 min at 430 °C for MgCl₂ vapour processing.

Current–voltage measurement. *J–V* analysis was performed under an AM1.5 spectrum at 1,000 W m^{−2} using a TS Space Systems solar simulator.

External quantum efficiency measurement. EQE measurements were made using a Bentham PVE300 system.

Capacitance–voltage measurements. *C–V* measurements were performed in the dark using a Solatron SI1260 impedance analyser and a frequency of 10 kHz. *C–V* data were recalculated into depth–density profiles using the method detailed in Blood and Orton²⁴. Only data recorded for bias <0.5 V were used, as at high forward bias the device back contact may dominate the capacitance response¹⁷.

SIMS. SIMS analysis was carried out by Loughborough Surface Analysis using a Cameca ims 4f instrument.

Back contact barrier height measurement. The formation of ohmic contacts to p-type CdTe is difficult owing to the high electron affinity of CdTe ($\chi_s = 4.5$ eV), meaning that a metal of work function >6.0 eV is required. Most metal contacts to CdTe therefore form a Schottky barrier at the back contact and this leads to the phenomenon of 'roll-over' (that is, a decrease in rate of current increase) at high forward bias. The back-contact barrier height may be determined from dark *J–V* measurements as a function of temperature: *J–V–T* measurement. Using the method proposed by Bätznert *et al.*¹⁴, the series resistance, R_s , is determined at forward bias above V_{oc} as a function of temperature. $R_s(T)$ may then be separated into an ohmic and an exponential component, which results from passage of the carriers over the back contact via thermionic emission. R_s may be expressed as follows:

$$R_s = R_{\Omega_0} + \frac{\partial R_{\Omega_0}}{\partial T} T + \frac{C}{T^2} e^{\left(\frac{\phi_b}{kT}\right)}$$

where R_{Ω_0} and $\frac{\partial R_{\Omega_0}}{\partial T}$ are the ohmic resistance and its temperature coefficient respectively. K is the Boltzmann constant and C is a fitting parameter. The barrier height ϕ_b may therefore be determined via exponential fitting of $R_s(T)$. *J–V–T* measurements were performed in a cryostat (CTI Cryogenics) using a temperature range of 150–350 K.

CdTe_{1–x}Mg_x as a potential electron back reflector. There has been no evidence reported to suggest Mg may act as an electrically active impurity centre in CdTe. The most likely form Mg may be expected to take in CdTe is via the formation of CdTe_{1–x}Mg_x alloyed phases. Rather than being problematic, CdTe_{1–x}Mg_x has in fact been investigated as a candidate for an electron back reflector layer for CdTe²⁵, owing to its excellent lattice matching to CdTe and expanded bandgap. The electron back reflector concept involves the incorporation of a wider bandgap material with a negligible valence band offset between the CdTe and back contact, thus forming a potential barrier in the device conduction band. This barrier reflects minority carrier electrons away from the back surface, thus reducing the back-surface recombination and improving the V_{oc} (ref. 26). As we have shown in this work, Mg may indeed be diffused into the CdTe without unduly comprising the CdTe doping profile and device performance, so CdTe_{1–x}Mg_x does have significant potential as an electron reflector layer. However, while it seems probable that some CdTe_{1–x}Mg_x phases may be present, as yet we are unable to find evidence of alloyed CdTe_{1–x}Mg_x layers being formed during MgCl₂ processing.

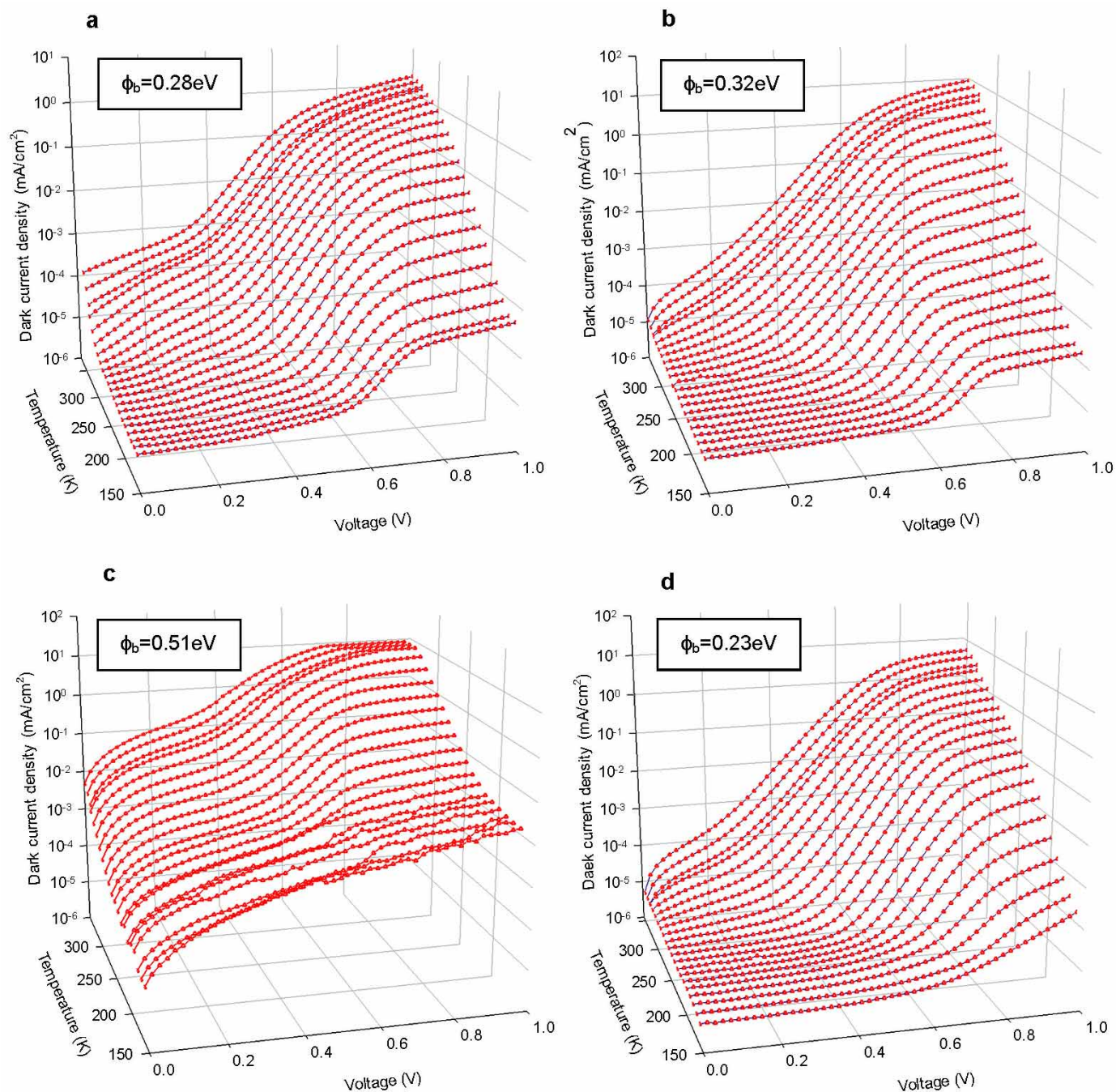
Additional MgCl₂ process development. Following the self-consistent comparative study of different Cl treatments, alterations to the cell structure and processing were made to improve the peak device performance attainable (see Extended Data Fig. 3). These changes to the cell fabrication process are listed below.

CdS:O nanostructured window layer. Nanostructured CdS:O window layers, which have an increased bandgap owing to quantum confinement effects²¹, were incorporated to replace the SnO₂/CdS buffer and window layer stack. Deposition was performed by radio-frequency sputtering using 60 W of power and at room temperature. A mixed argon/oxygen ambient gas was used with a 7% oxygen composition, giving a film with an as-deposited bandgap of about 3.9 eV and a film thickness of 250 nm. During CdTe deposition a portion of the CdS:O layer recrystallizes to form a thin (~40 nm) CdS layer, with a bandgap of 2.4 eV as the CdTe interface.

Aqueous MgCl₂ solution processing. Further development of the MgCl₂ post-growth treatment has led to MgCl₂ being deposited from a 1 M solution in deionized water. The MgCl₂/H₂O solution is spray deposited onto the back surface of the CdTe before annealing, allowing much finer control and uniformity. In addition, this negates any issues regarding the hygroscopic nature of MgCl₂. Optimal processing conditions were found to be 20 min annealing at 430 °C in air. This is now established as the preferred route for MgCl₂ deposition, as it requires the use of no solvents or thermal deposition equipment and has led to improved device efficiencies.

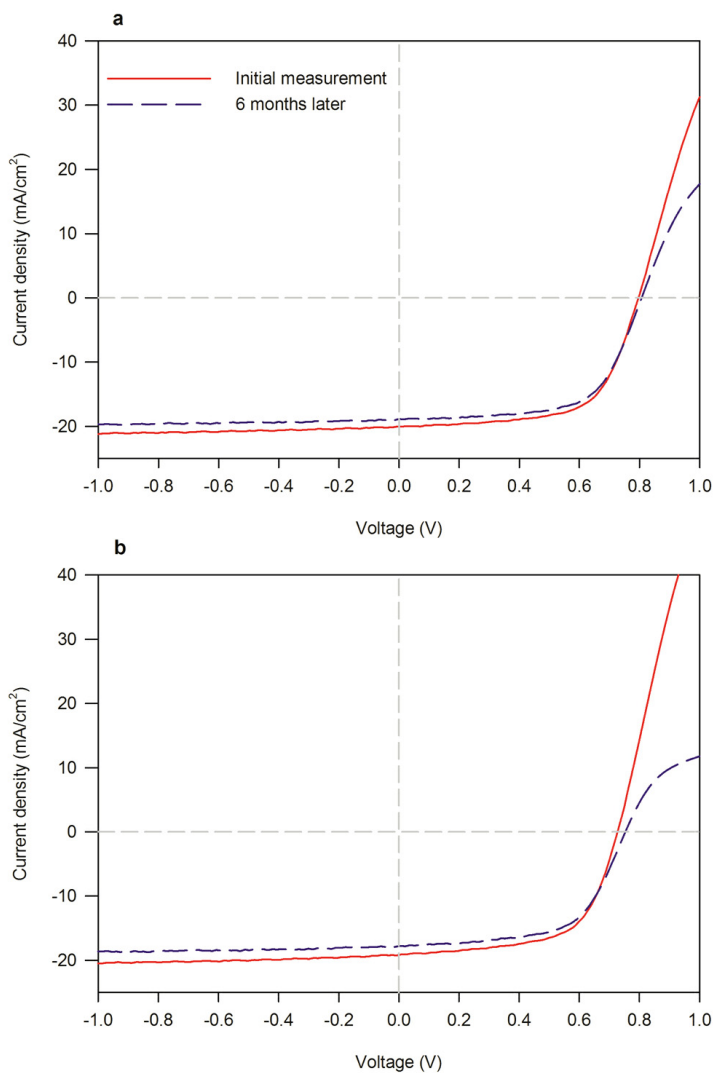
Cu back-contacting. Cu back-contacting is known to reduce the back-contact barrier in CdTe devices via the formation of a Cu_xTe_{1–x} phase at the CdTe free surface. Following post-growth activation treatment the cell is etched in nitric-phosphoric solution to form a Te-rich surface. A 2-nm-thick Cu film is then deposited via thermal evaporation before annealing at 150 °C for 20 min under vacuum. The back contact is then completed by the deposition of 60-nm of Au via thermal evaporation at room temperature.

23. Major, J. D., Proskuryakov, Y. Y. & Durose, K. Impact of CdTe surface composition on doping and device performance in close space sublimation deposited CdTe solar cells. *Prog. Photovolt. Res. Appl.* **21**, 436–443 (2013).
24. Blood, P. & Orton, J. *The Electrical Characterisation of Semiconductors: Majority Carriers and Electron States* 220–264 (Academic, 1992).
25. Kobayakov, P. S., Geisthardt, R., Cote, T. & Sampath, W. S. Growth and characterization of Cd_{1–x}Mg_xTe thin films for possible application in high-efficiency solar cells. In *Conference Record of the Twenty Sixth IEEE Photovoltaic Specialists Conference (Austin, Texas)* 160–163 (IEEE, 2012).
26. Hsiao, K.-J. & Sites, J. R. Electron reflector to enhance photovoltaic efficiency: application to thin-film CdTe solar cells. *Prog. Photovolt. Res. Appl.* **20**, 486–489 (2012).



Extended Data Figure 1 | J - V - T data for cells with various chloride treatments. Current density versus voltage curves measured as a function of temperature (J - V - T) with inset back-contact barrier height values ϕ_b , determined for the highest-efficiency contacts for: the CdCl_2 -treated device (a), the MgCl_2 -vapour-treated device (b), the NaCl -treated device (c) and the

high-efficiency cell (see Extended Data Fig. 3) treated with a 1 M $\text{MgCl}_2/\text{H}_2\text{O}$ solution and addition of 2 nm Cu to the back contact (d). Values for the back-contact barrier height are extracted by fitting to the temperature dependence of the series resistance R_s (see Methods).



c

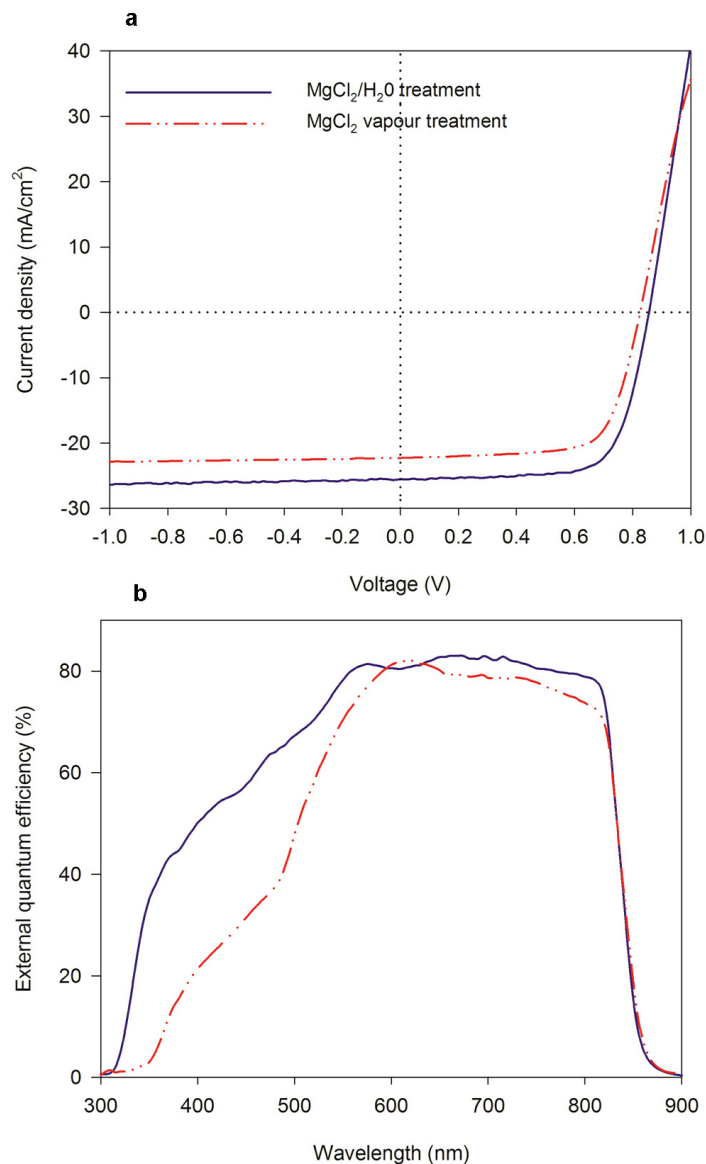
	Average performance ratio	Error
Efficiency	0.949	±0.032
Fill factor	0.997	±0.028
J _{sc}	0.945	±0.005
V _{oc}	1.008	±0.012

d

	Average performance ratio	Error
Efficiency	0.943	±0.011
Fill factor	0.987	±0.005
J _{sc}	0.929	±0.009
V _{oc}	0.987	±0.006

Extended Data Figure 2 | Stability measurements for CdCl₂- and MgCl₂-treated cells. *J*-*V* curves for devices treated with the MgCl₂ vapour process (a) and the CdCl₂ treatment (b). *J*-*V* curves were measured immediately after processing and then after 6 months of storage under ambient conditions. Performance degradation over the 6-month period was assessed from the

average shift in efficiency, fill factor, short circuit current density J_{sc} and open circuit voltage V_{oc} for nine contacts over this period. The averages for the ratio of initial and final performances, along with the associated error, are given for the MgCl₂ vapour treatment (c) and the CdCl₂ treatment (d).



Treatment	Efficiency (%)	Fill factor (%)	J _{sc} (mA/cm ²)	V _{oc} (V)
MgCl ₂ /H ₂ O	15.7	71.29	25.54	0.857
MgCl ₂ Vapour	13.5	70.24	23.26	0.826

Extended Data Figure 3 | High-efficiency MgCl₂-treated devices. Performance of CdTe devices treated with MgCl₂ is further improved following device optimization and the use of a 1 M MgCl₂/H₂O solution. A 2-nm Cu layer is added to the back contact, the ZnO buffer layer is replaced with a nanostructured CdS:O layer and CdS thickness is reduced to about 40 nm.

a, *J*–*V* curves for the unimproved MgCl₂-vapour-treated device (13.5%), and the improved cell treated with MgCl₂/H₂O solution (15.7%). **b**, EQE curves for the same devices, showing minimised CdS/ZnO cut-off at short wavelength (300–525 nm) by use of CdS:O layer. **c**, Extracted device performance parameters from *J*–*V* data.

Pathway from subducting slab to surface for melt and fluids beneath Mount Rainier

R Shane McGary¹, Rob L. Evans¹, Philip E. Wannamaker², Jimmy Elsenbeck¹ & Stéphane Rondenay³

Convergent margin volcanism originates with partial melting, primarily of the upper mantle, into which the subducting slab descends^{1,2}. Melting of this material can occur in one of two ways. The flow induced in the mantle by the slab can result in upwelling and melting through adiabatic decompression^{1,3}. Alternatively, fluids released from the descending slab through dehydration reactions can migrate into the hot mantle wedge, inducing melting by lowering the solidus temperature^{2,4}. The two mechanisms are not mutually exclusive¹. In either case, the buoyant melts make their way towards the surface to reside in the crust or to be extruded as lava. Here we use magnetotelluric data collected across the central state of Washington, USA, to image the complete pathway for the fluid–melt phase. By incorporating constraints from a collocated seismic study⁵ into the magnetotelluric inversion process, we obtain superior constraints on the fluids and melt in a subduction setting. Specifically, we are able to identify and connect fluid release at or near the top of the slab, migration of fluids into the overlying mantle wedge, melting in the wedge, and transport of the melt/fluid phase to a reservoir in the crust beneath Mt Rainier.

Despite important efforts to understand the production and transport of fluid and melt phases in subduction zones, a number of outstanding questions remain. Do fluids released from the slab rise vertically into the mantle wedge², or are ascent paths more complex^{6–8}? Do the fluids migrate into the mantle wedge by reactive porous flow^{1,9} or more quickly by way of fractures², channelling^{8,10} or diapirism³? Is the location of the volcanic arc defined by the location of melting above the anhydrous solidus¹¹, aqueous fluid connectivity in the mantle wedge¹², or some combination of kinematic variables such as slab dip and convergence rate with fluid transport¹³? Better constraints on the fluid transport pathways within the subduction setting are needed to address these questions.

The CAFE (Cascadia Array for Earthscope) experiment was designed to collect co-located seismic and magnetotelluric data from instrumentation deployed along a dense west–east transect across central Washington (Fig. 1). The seismic results have been addressed previously^{5,14}, and here the magnetotelluric results are presented. Data were collected at 60 wide-band and 20 long-period magnetotelluric stations, with generally good-quality responses over the period range from ~ 0.005 s to ~ 8000 s. The data are consistent with a two-dimensional, north–south-striking resistivity structure^{15,16}, which we determined through a two-dimensional nonlinear conjugate-gradient inversion¹⁷ (see Methods).

The magnetotelluric method is attractive for use in subduction settings because it is particularly sensitive to electrically conductive fluid and melt phases and can therefore be used to illuminate fluid processes and pathways in the subduction zone. The magnetotelluric method has been used to good effect in a number of subduction settings, including Cascadia^{18–21}. For the CAFE data, we were able to augment the magnetotelluric image by incorporating results from the seismic project directly into the magnetotelluric inversion process.

Results for both the seismic and magnetotelluric experiments are shown in Fig. 2. The velocity increase in the dipping low-velocity layer in the seismic image at depths of about ~ 40 – 45 km depth at the top of the

subducting layer is interpreted to reflect the transition of the hydrated basalts of the upper crust towards eclogite. The disappearance of that layer at depths of 75–90 km is further interpreted to result from the transition of lower-crustal metastable gabbro into eclogite. This reaction would be accelerated by fluids released from the dehydration of serpentine or chlorite in the subducting upper-mantle harzburgite²², a conclusion supported by local earthquake hypocentre evidence²³. The low-velocity feature above the subducting slab at depths of 65–80 km is interpreted as a fluid/melt phase resulting from the release of fluids from the slab⁵.

The magnetotelluric results are consistent with the seismic interpretation, but develop our understanding of the subduction process much further. The most prominent magnetotelluric structure is the highly conductive (2 – $5 \Omega \text{ m}$) region (A) near the top of the slab in the magnetotelluric model, coincident with the low-velocity fluid/melt feature in the seismic image. This conductor probably cannot be explained by dry melting alone as resistivities $< 5 \Omega \text{ m}$ would require excessively high melt fractions²⁴. Additionally, the temperatures near the top of the slab at this depth are roughly $1,100^\circ\text{C}$ – $1,150^\circ\text{C}$ (ref. 25) more than 200°C below the dry peridotite solidus¹. Both of these difficulties are resolved by the addition of water released from the slab. As little as 0.2 wt% water dissolved into the peridotite is sufficient to reduce the solidus temperature to below the temperatures found at the slab surface^{1,8}, which would allow flux melting to occur¹³. An incipient melt fraction of 2% would equate to 10 wt% water in the melt, enough to account for a resistivity of $2 \Omega \text{ m}$ (Fig. 3).

The calculations for fluid content assume that the fluid/melt phases are well connected. Interconnection along grain edges in a peridotite matrix requires a dihedral angle of $< 60^\circ$. At 25 kbar, this occurs at temperatures above 950°C (ref. 12), which is achieved in the region of interest. Although higher temperatures are required to maintain a sufficiently

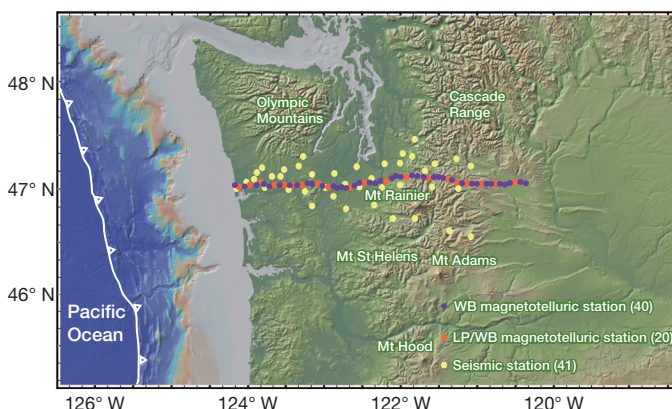


Figure 1 | Map showing station locations for the CAFE seismic and magnetotelluric stations (wideband and long-period) across central Washington state, USA. The numbers in parentheses indicate the number of stations for each category. WB, wideband; LP, long-period.

¹Department of Geology and Geophysics, MS#22, Woods Hole Oceanographic Institution, 360 Woods Hole Road, Woods Hole, Massachusetts 02543, USA. ²Energy and Geoscience Institute, University of Utah, 423 Wakara Way, Suite 300, Salt Lake City, Utah 84108, USA. ³Department of Earth Science, University of Bergen, Allégaten 41, 5007 Bergen, Norway.

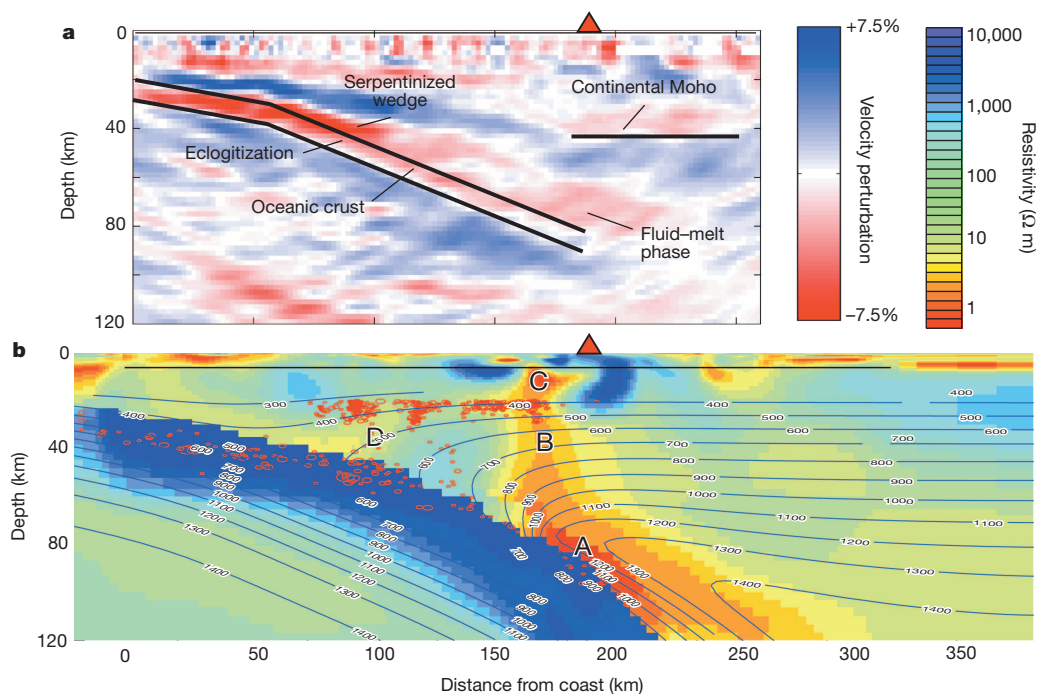


Figure 2 | Primary seismic (a) and magnetotelluric (b) models. Panel **b** includes both a thermal profile (contours, labelled in degrees Celsius) and earthquake hypocentre locations (red circles) within 20 km of our profile line²³. Fluid released from the subducting slab enters the mantle wedge at A. Melt initiated at or very near the interface is transported upward by buoyancy and dragged down. The fluid/melt phase rises through the mantle wedge (B) until it reaches the crust, joining fluids released from shallower reactions (D). The combined fluid/melt continues to rise until reaching a reservoir (C) in the crust. Mount Rainier is shown as a red triangle.

small dihedral angle at shallower depths, melt should become directly interconnected once the melt fraction exceeds 2 vol.% (ref. 26), ensuring that the fluid/melt phase remains well connected during ascent as the water-rich initial melt reacts with the overlying mantle⁴.

Our model cannot determine whether melting starts at the top of the slab or some short distance away, as the solidus for hydrated peridotite and the temperatures expected at the top of the slab are very close^{1,25}. However, yttrium concentrations in the range 13–19 parts per million (p.p.m.) in samples of Mt Rainier andesite suggest that some melting does occur at the interface and even within the slab for Cascadia^{27,28}.

In either case, the buoyant fluid and melt phases are gravitationally unstable and can rise through the mantle wedge diapirically³. It has also been argued that the trajectories for these instabilities may not necessarily be vertical, as the subduction induced motion in the mantle wedge would tend to drag them towards the hot corner of the wedge⁷. Both types of migration may be apparent in the magnetotelluric model (Fig. 2b), with the conductor extending from the primary source (A) both upwards into the mantle and sub-horizontally away from the slab. Additionally, the sub-horizontal extension could simply represent an extended zone of fluid release from the slab.

Although the incipient melt is very water-rich, it becomes diluted as it rises and further decompressional melting occurs⁸. This tradeoff would result in a slight increase in the bulk resistivity, for example, at 1,150 °C the resistivity of peridotite bearing a 2.5 vol.% melt with 10% water content is about 2 Ω m, whereas the resistivities as the melt fraction rises to 5 vol.% (5 wt% water content) and 10 vol.% (2.5 wt% water content) are about 3.5 Ω m and about 5 Ω m respectively (Fig. 3), ignoring the effects of water on bulk mantle. This is consistent with structure within the mantle wedge in the magnetotelluric model (feature B in Fig. 2b). This conductor maintains a resistivity of 5–6 Ω m throughout its ascent within the mantle. Figure 3 shows the effect of melt fraction and water content on the conductivity of the melt-bearing peridotite at 1,150 °C. Conductivities fall rapidly with decreasing temperature for a given combination of melt fraction and water content, suggesting that the resistivities seen in the image appear to rule out any significant temperature decrease (the thermal model superimposed on the image is a steady-state solution that does not factor in thermal energy transported by rising melt). This result argues for a relatively rapid vertical transit for the melt, possibly through large diapirs or interconnected conduits.

The ascending melt appears to rise until it reaches a reservoir in the crust (C), after being joined by fluids presumably originating from a conductor (D) associated with the dehydration of hydrated metabasalt in the upper-crustal layer of the descending slab⁵. Conductor C is in the same position as a conductor identified as the Southern Washington Cascades Conductor in earlier studies^{21,29}, and similar conductors appear to be a ubiquitous feature in subduction settings³⁰. It has been argued that this crustal reservoir represents metasediments²¹ or collected fluids³⁰, and this clearly seems to be the case in Oregon (and undoubtedly elsewhere) where the temperature of the fluid source is only about 500 °C. In the CAFE line, however, we can clearly see that the conditions for mantle melting are met. This does not rule out the possible presence of metasediments which may explain the small shallow conductor west of C, or even part of the conductive signature of C itself.

A conductor similar to D has also been identified previously in central Oregon²⁰. In the Oregon image, this conductor appeared to be the primary source of fluid connected to the crustal conductor, whereas in the CAFE line it is clearly a secondary source, with a much stronger contribution coming from the rising melt. Three-dimensional inversions

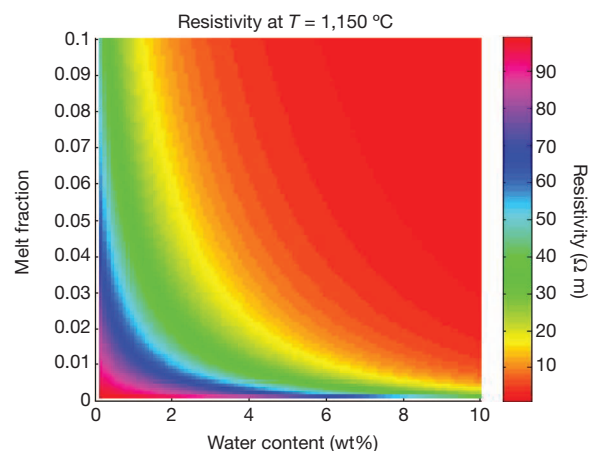


Figure 3 | The resistivity of peridotite for a given melt fraction and water content (within the melt) at a temperature of 1,150 °C. The method used for the calculation of resistivities is that given in ref. 24.

of regional magnetotelluric²¹ data at a coarser resolution than our data set corroborate our model by showing a strong conductor rising from the slab in this region, but also highlight the substantial along-strike variability in deep melt production in Cascadia. We suggest that this variability relates to first-order differences in the hydration of the incoming plate, with a wetter slab present beneath the CAFE line. Illuminating the differences in fluid release and melt generation brings us one step closer to understanding the connections between fluid release and seismogenic and volcanic processes operative in these critical tectonic settings.

METHODS SUMMARY

The CAFE magnetotelluric experiment comprised a dense linear array of 60 wide-band (spaced at about 5-km intervals) and 20 long-period (spaced at about 15-km intervals) magnetotelluric instruments. The former were acquired through contract from the University of Utah to Quantec Geoscience, Inc., while the latter were acquired by P.E.W. and student/post-doc crew using Narod Intelligent Magnetotelluric System (NIMS) instruments then owned by the University of Washington as part of the Electromagnetic Studies of the Continents (EMSOC) pool. The magnetotelluric array passes in an east–west line through the earlier CAFE seismic profile.

The long-period and wideband magnetotelluric data were processed using robust methods^{31,32} with two separate remote reference sites (one reference was over 500 km away). Combined, the instruments yielded magnetotelluric response functions typically over the period range of 0.005 s to 8000 s.

Multi-site, multi-frequency Groom–Bailey tensor decomposition^{15,16} was used to determine strike direction and viability of a two-dimensional inversion, and to separate from the impedance tensor distortion elements affecting both amplitude and phase of the electric field.

We generated our two-dimensional models of resistivity structure using the WinGLink inversion algorithm¹⁷. The inversions started with a 100 Ω m half-space, excepting the ocean, fixed at 0.33 Ω m and defined by local bathymetry, and a dipping resistor representing the upper part of the subducting slab whose location was defined by constraints from the seismic migration results. A tear zone (allowing sharp conductivity transitions) was imposed at the top of the dipping resistor. Extensive testing for sensitivity and robustness was performed on significant features by removing features from the model and re-running the inversion, and by varying parameters such as resistivity or extent of a feature in a systematic way and observing the effect on the misfit between the data and model. The incorporation of seismic constraints into the magnetotelluric inversion constitutes a novel approach and enables superior imaging of fluids in the subduction setting.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 December 2013; accepted 15 May 2014.

1. Stern, R. J. Subduction zones. *Rev. Geophys.* **40**, 1012, <http://dx.doi.org/10.1029/2001RG000108> (2002).
2. Davies, J. H. & Stevenson, D. J. Physical model of source region of subduction zone volcanics. *J. Geophys. Res.* **97** (B2), 2037–2070 (1992).
3. Hall, P. S. & Kincaid, C. Diapiric flow at subduction zones: a recipe for rapid transport. *Science* **292**, 2472–2475 (2001).
4. Grove, T. L., Chatterjee, N., Parman, S. W. & Medard, E. The influence of H₂O on mantle wedge melting. *Earth Planet. Sci. Lett.* **249**, 74–89 (2006).
5. McGary, R. S. The CAFE experiment: a joint seismic and MT investigation of the Cascadia subduction system. PhD thesis, MIT/WHOI Joint Program (2013).
6. Schurr, B., Asch, G., Rietbrock, A., Trumbull, R. & Haberland, C. Complex patterns of fluid and melt transport in the central Andean subduction zone revealed by attenuation tomography. *Earth Planet. Sci. Lett.* **215**, 105–119 (2003).
7. Marschall, H. R. & Schumacher, J. C. Arc magmas sourced from mélange diapirs in subduction zones. *Nature Geosci.* **5**, 862–867 (2012).
8. Pearce, J. A. & Peate, D. W. Tectonic implications of the composition of volcanic arc magmas. *Annu. Rev. Earth Planet. Sci.* **23**, 251–285 (1995).
9. Navon, O. & Stolper, E. Geochemical consequences of melt percolation: the upper mantle as a chromatographic column. *J. Geol.* **95**, 285–307 (1987).
10. Daines, M. J. & Kohlstedt, D. L. The transition from porous to channelized flow due to melt/rock reaction during melt migration. *Geophys. Res. Lett.* **21**, 145–148 (1994).
11. England, P. C. & Katz, R. F. Melting above the anhydrous solidus controls the location of volcanic arcs. *Nature* **467**, 700–703 (2010).

12. Mibe, K., Yoshino, T., Ono, S., Yasuda, A. & Fujii, T. Connectivity of aqueous fluid in eclogite and its implications for fluid migration in the Earth's interior. *J. Geophys. Res.* **108** (B6), 2295, <http://dx.doi.org/10.1029/2002JB001960> (2003).
13. Grove, T. L., Till, C. B., Lev, E., Chatterjee, N. & Medard, E. Kinematic variables and water transport control the formation and location of arc volcanoes. *Nature* **459**, 694–697 (2009); erratum **460**, 1044 (2009).
14. Abers, G. A. *et al.* Imaging the source region of Cascadia tremor and intermediate-depth earthquakes. *Geology* **37**, 1119–1122 (2009).
15. McNeice, G. W. & Jones, A. G. Multisite, multifrequency tensor decompositions of magnetotelluric data. *Geophysics* **66**, 158–173 (2001).
16. Groom, R. W. & Bailey, R. C. Decomposition of magnetotelluric impedance tensors in the presence of local three-dimensional galvanic distortion. *J. Geophys. Res.* **94**, 1913–1925 (1989).
17. Rodi, W. L. & Mackie, R. L. Nonlinear conjugate gradients algorithm for 2-D magnetotelluric inversion. *Geophysics* **66**, 174–187 (2001).
18. Wannamaker, P. E. *et al.* Resistivity cross section through the Juan de Fuca subduction system and its tectonic implications. *J. Geophys. Res.* **94**, 14127–14144 (1989).
19. Soyer, W. & Unsworth, M. Deep electrical structure of the northern Cascadia (British Columbia, Canada) subduction zone: implications for the distribution of fluids. *Geology* **34**, 53–56 (2006).
20. Evans, R. L., Wannamaker, P., McGary, R. S. & Elsenbeck, J. Electrical structure of the Central Cascadia subduction zone: the EMSLAB Lincoln line revisited. *Earth Planet. Sci. Lett.* <http://dx.doi.org/10.1016/j.epsl.2013.04.021> (in the press).
21. Meqbel, N. M., Egbert, G. D., Wannamaker, P. E., Kelbert, A. & Schultz, A. Deep electrical resistivity structure of the northwestern U.S. derived from 3-D inversion of USArray magnetotelluric data. *Earth Planet. Sci. Lett.* <http://dx.doi.org/10.1016/j.epsl.2013.12.026> (in the press).
22. John, T. & Schenk, V. Partial eclogitization of gabbroic rocks in a late Precambrian subduction zone (Zambia): prograde metamorphism triggered by fluid infiltration. *Contrib. Mineral. Petrol.* **146**, 174–191 (2003).
23. McCrory, P. A., Blair, J. L., Waldhauser, F. & Oppenheimer, D. H. Juan de Fuca slab geometry and its relation to Wadati-Benioff zone seismicity. *J. Geophys. Res.* **117**, B09306 (2012).
24. Ni, H., Keppler, H. & Behrens, H. Electrical conductivity of hydrous basaltic melts: implications for partial melting in the upper mantle. *Contrib. Mineral. Petrol.* **162**, 637–650 (2011).
25. van Keken, P. E., Hacker, B. R., Syracuse, E. M. & Abers, G. A. Subduction factory: 4. Depth-dependent flux of H₂O from subducting slabs worldwide. *J. Geophys. Res.* **116**, B01401 (2011).
26. Zhu, W., Gaetani, G. A., Fusses, F., Montesi, L. G. J. & De Carlo, F. Microtomography of partially molten rocks: three-dimensional melt distribution in mantle peridotite. *Science* **332**, 88–91 (2011).
27. Defant, M. J. & Drummond, M. S. Derivation of some modern arc magmas by melting of young subducted lithosphere. *Nature* **347**, 662–665 (1990).
28. Stockstill, K. R., Vogel, T. A. & Sisson, T. W. Origin and emplacement of the andesite of Burroughs Mountain, a zoned, large-volume lava flow at Mount Rainier, Washington, USA. *J. Volcanol. Geotherm. Res.* **119**, 275–296 (2003).
29. Egbert, G. D. & Booker, J. R. Imaging crustal structure in southwestern Washington with small magnetometer arrays. *J. Geophys. Res.* **98** (B9), 15967–15985 (1993).
30. Worzewski, T., Jegen, M., Kopp, H., Brasse, H. & Castillo, W. T. Magnetotelluric image of the fluid cycle in the Costa Rican subduction zone. *Nature Geosci.* **4**, 108–111 (2011).
31. Chave, A. D. & Thompson, D. J. Bounded influence magnetotelluric response function estimation. *Geophys. J. Int.* **157**, 988–1006 (2004).
32. Egbert, G. D. Robust multiple station magnetotelluric data processing. *Geophys. J. Int.* **130**, 475–496 (1997).

Acknowledgements We thank V. Maris, M. Brown, A. Kelbert, and Quantec Geoscience, Inc. for their part in data acquisition. We also thank A. Pommier, D. Lizarralde, A. Malcolm, A. Shaw, H. Marschall and J. P. Canales for critical discussions and input on early versions of the manuscript. Finally, we thank P. van Keken for use of his thermal overlay in the magnetotelluric figure. This work was supported by US National Science Foundation grant EAR08-44041 (Principal Investigator R.L.E.) and US National Science Foundation grant EAR08-43725 (Principal Investigator P.E.W.), both through the Earthscope programme. R.S.M. was supported by a National Defense Science and Engineering Graduate (NDSEG) fellowship.

Author Contributions R.L.E. and P.E.W. conceived the experiment. R.S.M. participated in data collection and was primarily responsible for the processing and inversion work and analysis. R.L.E. was involved in all aspects of the development of the magnetotelluric models. P.E.W. coordinated and led the data collection, and also performed some of the processing and analysis of the broadband data. J.E. assisted with data reduction and processing. S.R. was involved in the production of the seismic image. All authors contributed to the understanding of the results and editing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.S.M. (mcgaryr@tcnj.edu).

METHODS

Time series data for the CAFE magnetotelluric project were collected at 20 long-period and 60 wideband sites. The long-period data were collected using NIMS instruments with each station in place for typically three weeks. The wideband data were collected using Phoenix Geosystems instruments, with a typical recording interval of 15 h for each site. The entire array was 280 km in length, and designed to be roughly collocated with the earlier CAFE passive seismic experiment.

The time series data from each long-period station were visually inspected for breaks, trends, and signal-to-noise ratio, and then windowed with data from two separate magnetic remote reference sites. One remote reference for each station was chosen from distant CAFE magnetotelluric stations, and the second concatenated from Earthscope/USArray stations that were operated simultaneously in Nevada and California, with a minimum distance of 500 km between stations to ensure that the noise between the stations was not likely to be correlated to any significant degree. Dual remote references were also used for the wideband data, with one remote reference located approximately 30 km east of Mt Rainier, and the other in Buena Vista Valley, Nevada.

The long-period time series data were processed into impedance tensors using the robust bounded influence remote reference processing (BIRRP) algorithm³¹. The wideband data were also processed using robust methods³². The long-period transfer functions provided useful data from ten seconds up to several thousand seconds, and the wideband transfer functions provided useful data from less than one second up to several hundred.

Dimensionality and regional strike direction were evaluated for the data in a variety of ways, including phase tensor analysis³³, Bahr skew analysis³⁴, and multi-site, multi-frequency Groom–Bailey tensor decomposition using the Strike software package^{15,16}. The phase tensor ellipses show a consistent pattern for periods between 10 s and 2,000 s, with a near north–south strike for stations west of Mt Rainier, shifting to a slightly clockwise strike to the east of Mt Rainier, with a more pronounced clockwise shift for stations towards the eastern end of the profile. This is consistent with previous strike analysis conducted on a set of Earthscope long-period stations along a line at roughly 46.5° N (ref. 35).

Bahr (phase-sensitive) skew analysis provides a justification for a two-dimensional analysis of the data between the same periods, with exceptions beneath stations LP-32 and LP-36 (immediately west of Mt Rainier) for periods greater than 400 s, and beneath stations LP-44 (just east of Mt Rainier) and LP-54 (in a river valley just north of Ellensburg) for periods less than 150 s.

Using the Strike algorithm¹⁵, we determined that a regional strike direction of seven degrees west of north provided the best χ -squared fit for the long-period data for periods of 100–1,350 s, within confidence intervals when stations LP-32 and LP-36 were excluded. A strike direction of due north also produced an acceptable fit, and was used in the inversions. The strike angles for the entire set of decomposed transfer functions are displayed in Extended Data Fig. 1.

To generate our regularized two-dimensional models of resistivity structure, we used the nonlinear conjugate gradient inversion algorithm WinGLink¹⁷. Our inversion mesh consisted of 116 rows and 234 columns, with a very fine row height near the surface increasing gradually with depth, and a column width that was maintained to be generally uniform to the extent allowed by station spacing. One set of inversions were run starting with a half-space of 100 Ω m, except for the ocean, which was fixed at 0.33 Ω m with an extent defined by local bathymetry. A second set of inversions additionally incorporated a tear zone at the top of the slab, with the location determined by the seismic migration profile, and an imposed resistive feature (5,000 Ω m) that represents the slab, extending approximately 30 km below the slab surface. Although this feature was imposed on the incipient inversion model, it was not locked and was free to evolve during the inversion process. It has been shown that the tear zone and imposition of this resistive feature will produce a significantly more accurate image of nearby conductive fluids, particularly if these

fluids are released at or near the slab surface⁵. The inversion model generated without the imposed tear is displayed as Extended Data Fig. 2. The same technique has been used to invert data in Cascadia and other subduction zones^{20,36}.

The τ value determines the tradeoff between smoothness and misfit in the inversion, and we determined a τ value of 3.3 to be optimal. Two other parameters, α and β , define the relationship of the smoothness parameter in the vertical as compared to the horizontal, and the way in which smoothness changes with depth respectively. A range of values from 0.8 to 1.8 for α and 1.0 to 2.0 for β were evaluated, with the final selected values of 1.5 and 1.7 respectively. These values are consistent with values used in inversions of other similar data sets²⁰. While small changes in α and β have been shown to produce striking differences in structure in some cases³⁶, the changes that we saw in structure when varying these parameters was quite minimal. It should also be noted that the same parameter values were used in generating both final images, that is, with and without the tear imposed.

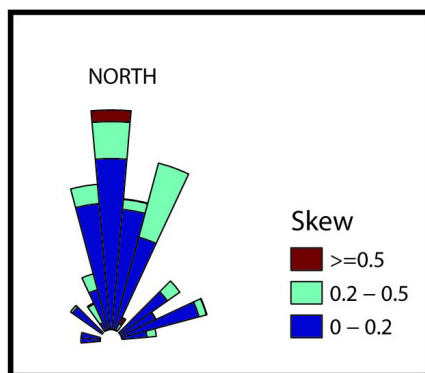
We generated models that inverted both the transverse magnetic (TM) and transverse electric (TE) modes, with error floors of 5% and 15% respectively, and a set of models using only the TM mode. The tipper function was included in both sets of models, and was also evaluated separately, clearly supporting the presence of the vertical conductor.

We also evaluated the extent to which the primary structures in our models were supported and/or required by the data. This was done in a number of ways, such as assessing how the structures were affected by parameter modifications as described above, and comparing the resulting models to the data pseudo-sections (included as Extended Data Fig. 3) to give us a better understanding of how data from each station might be affecting the inversion.

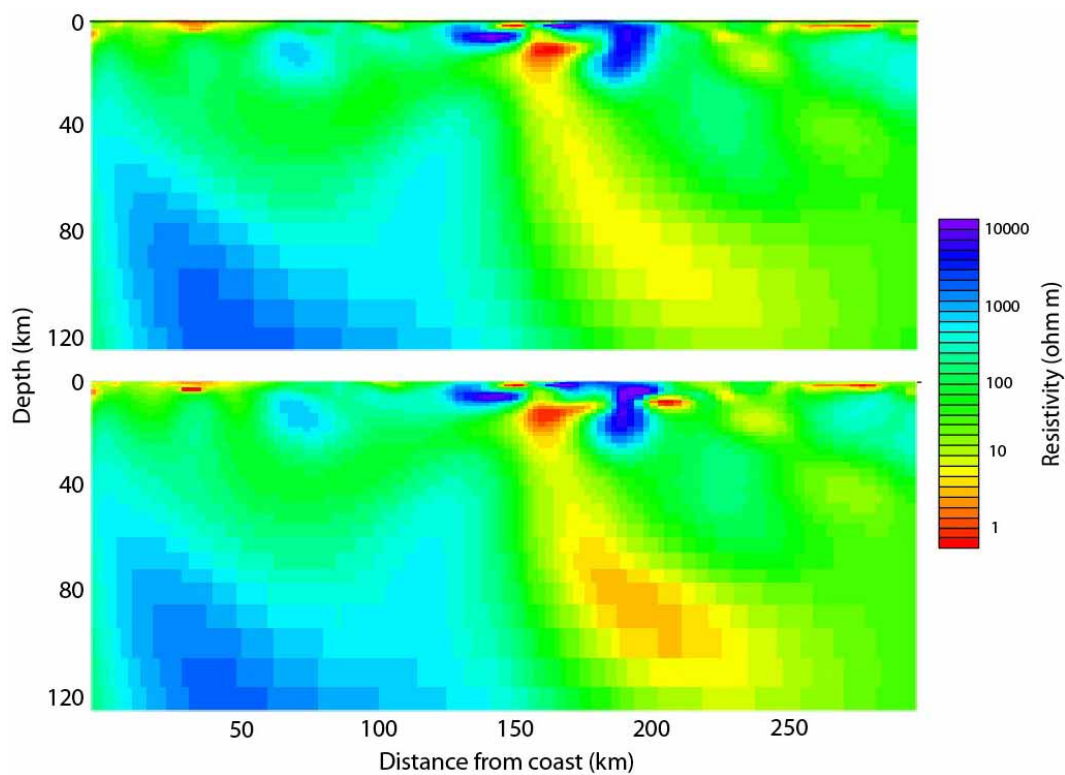
Additionally, we used manual editing techniques. This involved removing or altering a feature in the resultant model and then allowing the inversion process to seek a solution optimally close to this altered model. If the feature was restored by the inversion, it was taken to be required by the data. If the inversion was able to achieve a misfit comparable to the original misfit without restoring the feature, the structure in question was determined to be allowed but not required by the data. For our final models, all three major conductive features, the slab top, the vertical column, and the crustal conductor were found to be required by the data using these methods.

We achieved a root-mean-square misfit value of 1.89 for the inversion using the tear, as compared to 3.08 for the halfspace. A plot of root-mean-square misfit by station (Extended Data Fig. 4) shows that while the model with the tear shows improvement in fit for almost every station, the effect is much more pronounced between stations 30 and 55, which correspond to horizontal distances of 150–275 km from the coast. By far the most striking differences between the two models in this range is that for the model with the tear, the conductor is more intense and not separated from the resistive slab by any significant distance. Taken together, this strongly suggests that the model with the tear is significantly more accurate, and that the imposed smoothness is largely responsible for the higher misfit in the half-space model. We note that although the geometry of this data set is not optimized for a three-dimensional treatment, inversions of the regional Earthscope Transportable Array data collected on a ~70-km grid, confirms the presence of the large conductor emanating from the slab³⁵, although the resolution of the three-dimensional model is necessarily coarser than that in our model.

33. Caldwell, T. G., Bibby, H. M. & Brown, C. The magnetotelluric phase tensor. *Geophys. J. Int.* **158**, 457–469 (2004).
34. Simpson, F. & Bahr, K. *Practical Magnetotellurics* 93–98 (Cambridge Univ. Press, 2005).
35. Patro, P. K. & Egbert, G. D. Regional conductivity structure of Cascadia: preliminary results from 3D inversion of USArray transportable array magnetotelluric data. *Geophys. Res. Lett.* **35**, L20311 (2008).
36. Matsuno, T. et al. Upper mantle electrical resistivity structure beneath the central Mariana subduction system. *Geochem. Geophys. Geosyst.* **11**, Q09003 (2010).

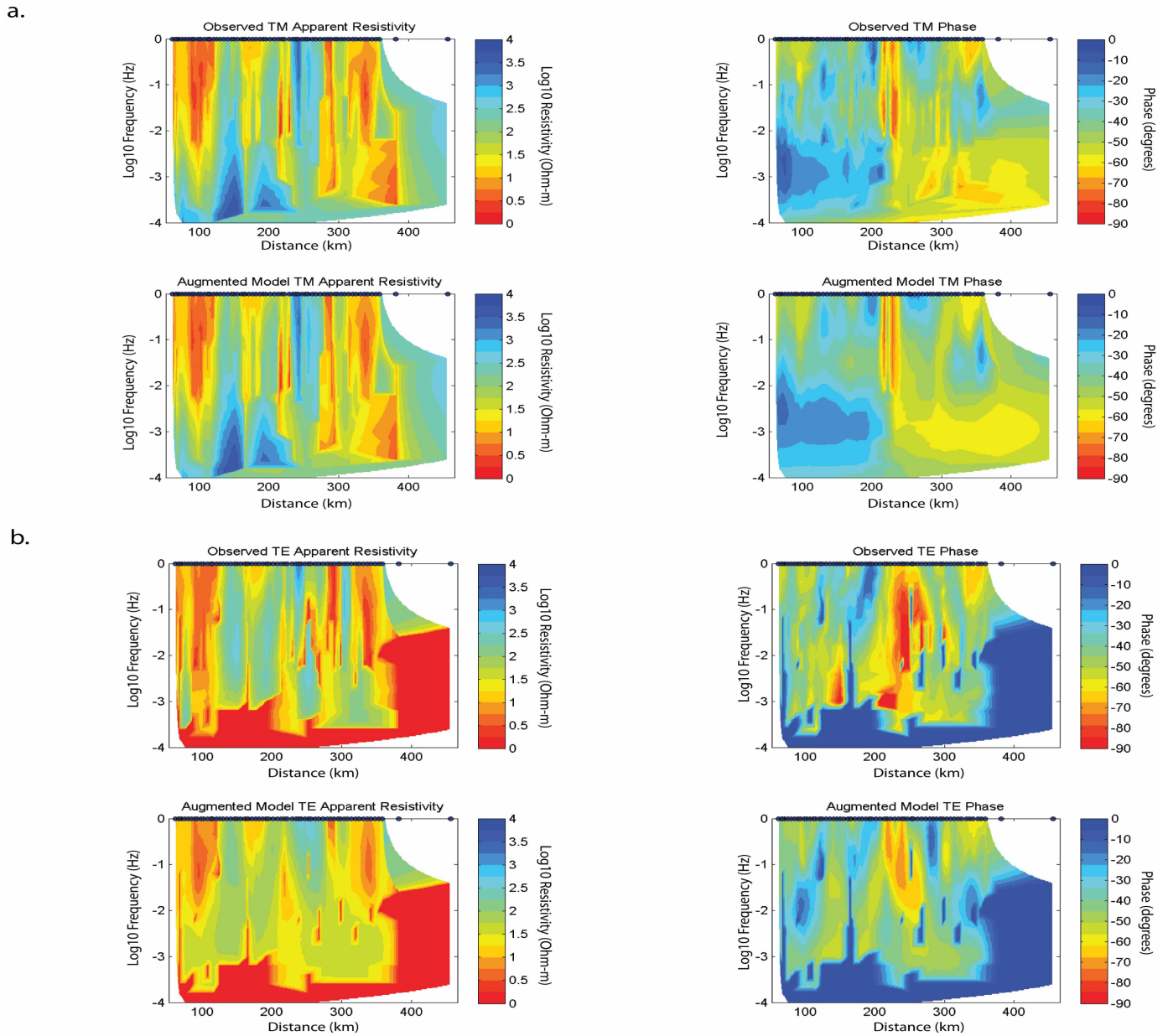


Extended Data Figure 1 | Rose diagram showing overall strike directions. The colour code reflects the Bahr skew as determined using the STRIKE algorithm¹⁵ for the CAFE data set.



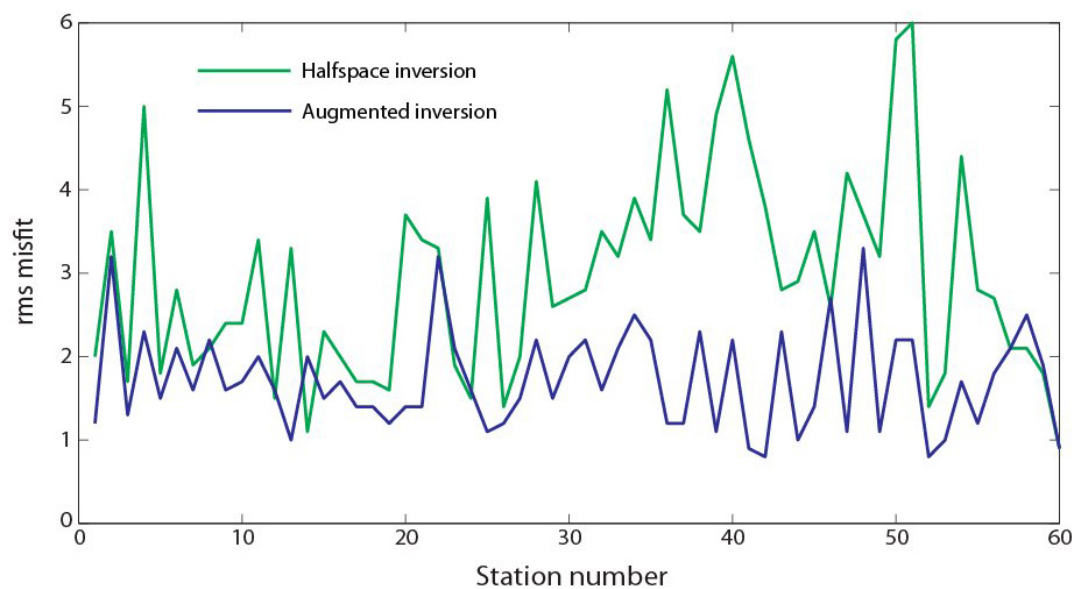
Extended Data Figure 2 | Primary standard inversion images for the CAFE data. These magnetotelluric images were generated without incorporating a tear zone on top of the slab or setting the initial resistivity for the slab. The top

image was generated using a combination of the TM mode and tipper, whereas the bottom image was produced using the TM and TE modes along with the tipper.



Extended Data Figure 3 | The TM (a) and TE (b) pseudo-sections for the CAFE magnetotelluric data. The two upper panels in **a** and the two upper panels in **b** show apparent resistivity and phase for the data. The two lower

panels in **a** and the two lower panels in **b** show apparent resistivity and phase for the model. Both models are limited horizontally to correspond with the surface covered by the CAFE magnetotelluric array.



Extended Data Figure 4 | Plot of root-mean-square misfit against the 60 CAFE magnetotelluric stations for the TM/TE/tipper models. The blue line shows root-mean-square misfit by station for the halfspace model (without a tear or initial resistivity set for the upper slab), and the green line shows the

same for the augmented model (using a tear at the top of the slab and imposing initial resistivity for the upper part of the slab). The overall root-mean-square misfit values were 3.08 for the halfspace model, and 1.89 for the augmented model.

Declines in insectivorous birds are associated with high neonicotinoid concentrations

Caspar A. Hallmann^{1,2}, Ruud P. B. Foppen^{2,3}, Chris A. M. van Turnhout², Hans de Kroon¹ & Eelke Jongejans¹

Recent studies have shown that neonicotinoid insecticides have adverse effects on non-target invertebrate species^{1–6}. Invertebrates constitute a substantial part of the diet of many bird species during the breeding season and are indispensable for raising offspring⁷. We investigated the hypothesis that the most widely used neonicotinoid insecticide, imidacloprid, has a negative impact on insectivorous bird populations. Here we show that, in the Netherlands, local population trends were significantly more negative in areas with higher surface-water concentrations of imidacloprid. At imidacloprid concentrations of more than 20 nanograms per litre, bird populations tended to decline by 3.5 per cent on average annually. Additional analyses revealed that this spatial pattern of decline appeared only after the introduction of imidacloprid to the Netherlands, in the mid-1990s. We further show that the recent negative relationship remains after correcting for spatial differences in land-use changes that are known to affect bird populations in farmland. Our results suggest that the impact of neonicotinoids on the natural environment is even more substantial than has recently been reported and is reminiscent of the effects of persistent insecticides in the past. Future legislation should take into account the potential cascading effects of neonicotinoids on ecosystems.

Although concerns have been raised about the direct effects of neonicotinoids on non-target vertebrate species⁸, neonicotinoids are in general thought to be less harmful to mammals and birds than to insects. The main mode of action of neonicotinoids occurs through binding nicotinic acetylcholine receptors in the central nervous system of invertebrates⁹, and neonicotinoids bind with substantially less affinity to these receptors in vertebrates¹⁰. This property has made neonicotinoids highly favoured agrochemicals worldwide over the past two decades¹¹. In the Netherlands, imidacloprid was first administered by the Board for the Authorisation of Plant Protection Products and Biocides (Ctgb) in August 1994. Annual use increased rapidly from 668 kg in 1995 to 5,473 kg in 2000 and 6,332 kg in 2004 (ref. 12). Since 2003, imidacloprid has ranked consistently in the top three pesticides that exceed the environmental concentrations permitted by quality standards in the Netherlands^{4,13}.

As neonicotinoids have relatively long half-lives in soil and are water soluble, they have the potential to accumulate in soils and to leach into surface water and ground water. Their systemic property (that is, their ability to spread through all of the tissues of the plants under treatment), together with their widespread use, indicates that many organisms in agricultural environments are likely to become exposed⁸. Indeed, studies have shown, both in experimental and in field conditions, that neonicotinoids may affect non-target invertebrate species across terrestrial and aquatic ecosystems^{4–6}. The question remains, however, whether the effects are sufficiently severe to affect ecosystems through trophic interactions: that is, beyond the direct lethal and sublethal effects on individual species. In the past, the introduction of insecticides has caused prey-base collapses, which in turn affected avian populations^{14–16}, showing that pesticide-induced declines in invertebrate densities can cause food deprivation for birds. Thus, if natural insect communities are indeed affected by neonicotinoids to the extent of causing disruptions in the food chain, we may expect insectivorous bird species to be affected as well.

The present study takes advantage of two standardized, long-term, country-wide monitoring schemes in the Netherlands (see Methods)—the Dutch Common Breeding Bird Monitoring Scheme¹⁷ and surface-water quality measurements⁴—to investigate the extent to which average concentrations of imidacloprid residues in the period 2003–2009 spatially correlate with bird population trends in the period 2003–2010. We selected 15 passerine species that are common in farmlands and depend on invertebrates during the breeding season (Extended Data Table 1 and Supplementary Methods). We interpolated concentrations of imidacloprid in surface water to bird monitoring plots (Extended Data Figs 1–3, Supplementary Data and Supplementary Methods) and examined how local bird trends correlate with these concentrations (Fig. 1).

The average intrinsic rate of increase in local farmland bird populations was negatively affected by the concentration of imidacloprid (Fig. 1b, linear mixed effects regression (LMER): d.f. = 1,443, $t = -5.64$, $P < 0.0001$). At the separately tested individual species level, 14 out of 15 of the tested species had a negative response to interpolated imidacloprid concentrations, and 6 out of 15 had a significant negative response at the 95% confidence level after Bonferroni correction (Table 1 and Extended Data Fig. 4). Thus, higher concentrations of imidacloprid in surface water in the Netherlands are consistently associated with lower or negative population growth rates of passerine insectivorous bird populations. From our analysis, the imidacloprid concentration above which bird populations were in decline was $19.43 \pm 0.03 \text{ ng l}^{-1}$ (mean \pm s.e.m.) (Fig. 1b). In areas with imidacloprid measurements above this concentration, bird populations declined by 3.5% on average annually.

We checked whether two alternative explanations could have caused spurious correlations between imidacloprid concentrations and bird population trends over the period 2003–2010. First, it is possible that our results could simply reflect a spatial pattern of local farmland bird declines that started before the introduction of imidacloprid¹⁸. Therefore, we tested whether declines were present before the introduction of imidacloprid, in 1994. In contrast to the strongly negative relationship between imidacloprid concentration and bird population trends in 2003–2010 (Fig. 1b), the 2003–2009 imidacloprid concentrations were not significantly associated with bird trends in the period 1984–1995 ($t = -1.43$, $P = 0.15$ for LMER_{<1995}; $t = -2.16$, $P = 0.031$ for LMER_{>2003}; using plots only with trend data for both periods, d.f. = 365; see Extended Data Fig. 6 and Supplementary Methods). Overall, bird population trends in these two periods, paired by plot and species, were uncorrelated ($r = -0.028$, Pearson product moment test; $t = -0.5455$, d.f. = 379, $P = 0.56$). We can thus conclude that the spatial pattern observed does not reflect long-term ongoing local declines caused by other factors. This finding suggests that imidacloprid is likely to have contributed to the declining population trend of the local birds.

Second, we tested whether spatial differences in land-use changes related to agricultural intensification confounded the effects of imidacloprid in our analyses. We performed multiple mixed effects regression analyses in which we included the local changes in land area use (urban area, natural area, and the production areas of maize, winter cereals and fallow land) and the amount of fertilizer applied (nitrogen in kg ha^{-1}) as fixed

¹Radboud University, Institute of Water and Wetland Research, Departments of Experimental Plant Ecology & Animal Ecology and Ecophysiology, PO Box 9100 (Mail Box 31), 6500 GL Nijmegen, The Netherlands. ²Sovon, Dutch Centre for Field Ornithology, PO Box 6521, 6503 GA Nijmegen, The Netherlands. ³Birdlife Netherlands, PO Box 925, 3700 AX Zeist, The Netherlands.

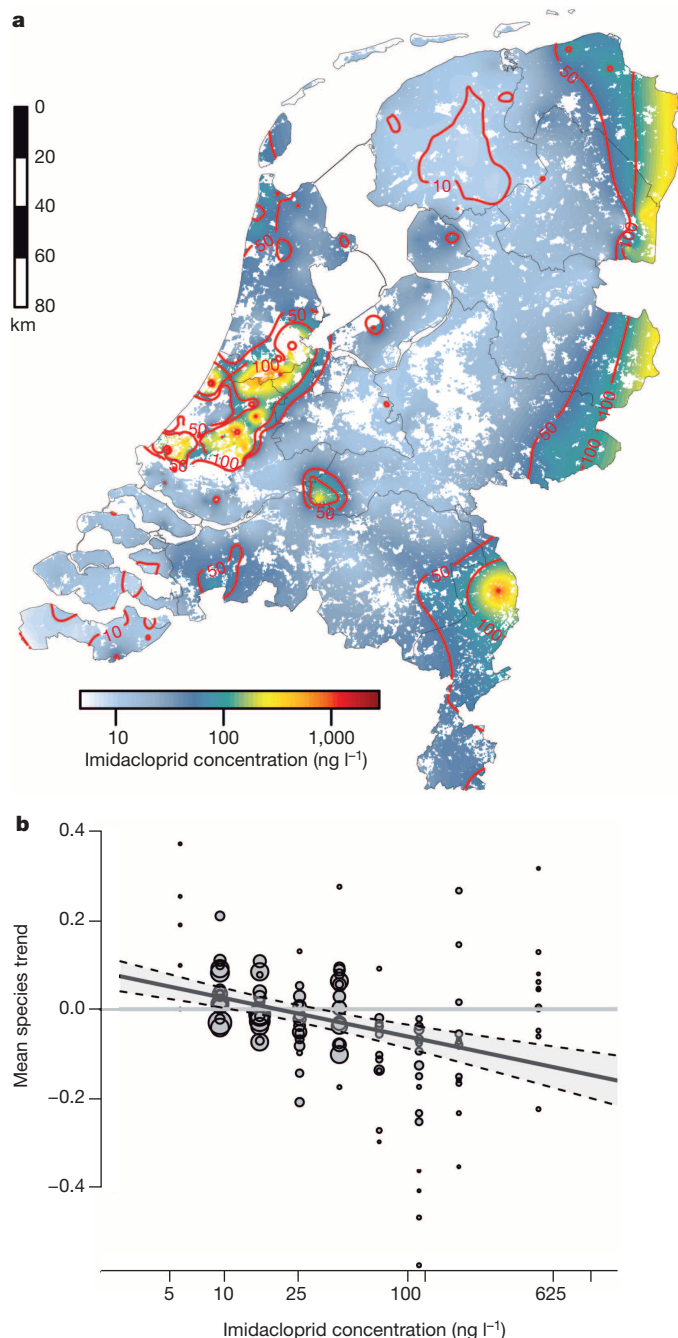


Figure 1 | Effect of imidacloprid on bird trends in the Netherlands.

a, Interpolated (universal kriging) mean logarithmic concentrations of imidacloprid in the Netherlands (2003–2009). **b**, Relationship between the average annual intrinsic rate of population increase over 15 passerine bird species and imidacloprid concentrations in Dutch surface water. Each point represents the average intrinsic rate of increase of a species over all plots in the same concentration class, whereas the size of the point is scaled proportionally to the number of species–plot combinations on which the calculated mean is based. Binning into classes was performed to reduce scatter noise and aid in visual interpretation. Actual analysis, and the depicted regression line, was performed on raw data ($n = 1,459$). The regression line is given by $0.1110 - 0.0374 \text{ (s.e.m.)} \times \log[\text{imidacloprid}]$ ($P < 0.0001$). Dashed lines delineate the 95% confidence interval.

explanatory variables (see Supplementary Data), in addition to imidacloprid concentrations. These variables have been put forward frequently as causal factors related to farmland bird declines^{19–21}, although their major effect may have already occurred earlier in the twentieth century. As imidacloprid usage is likely to be related to horticulture and greenhouses⁴, spatial changes in these variables may confound the effects of imidacloprid on bird trends. We therefore also incorporated changes in the area of greenhouses and the area of flower bulb production in our analysis. The results indicate that the concentration of imidacloprid and the changes in urban and natural areas were negatively correlated with local population trends, whereas the changes in the bulb and fallow land were positively correlated (Fig. 2). However, only imidacloprid and bulb area were significantly correlated with local trends (Extended Data Table 2).

So far, the suggested potential risks of neonicotinoids for birds have focused on the acute toxic effects caused by direct consumption⁸. Our results suggest another possibility: that is, that the depletion of insect food resources has caused the observed relationships. Two lines of evidence seem to support this. First, 9 out of 15 species tested in the present study are exclusively insectivorous. All 15 species feed their young (almost) exclusively with invertebrates, and food demand is the highest in this period. Adult skylarks, tree sparrows, common starlings, yellowhammers, meadow pipits and mistle thrushes are also granivorous to some extent and may thus directly consume coated seed. However, meadow pipits and mistle thrushes forage on seeds only outside the breeding season, and for all 15 species the bulk of the diet during the breeding season consists of invertebrates⁷. Second, recent *in situ* research involving the same areas as the present study revealed strong declines in insect macrofauna, including species that have a larval stage in water, where imidacloprid concentrations were elevated⁴. These insects (particularly Diptera, Ephemeroptera, Odonata, Coleoptera and Hemiptera) are an important food source in the breeding season for the bird species that we investigated⁷. However, as our results are correlative, we cannot exclude other trophic or direct ways in which imidacloprid may have an effect on the bird population trends. Food resource depletion may not be the only or even the most important cause of decline. Other possible causes of decline include trophic accumulation of this neonicotinoid through

Table 1 | Effect of imidacloprid on insectivorous bird species population trends

Species	Effect (mean)	Error (s.e.m.)	t value	P	n
Marsh warbler (<i>Acrocephalus palustris</i>)	0.0110	0.0187	0.5871	0.5584	105
Sedge warbler (<i>Acrocephalus schoenobaenus</i>)	−0.0229	0.0152	−1.5070	0.1351	99
Reed warbler (<i>Acrocephalus scirpaceus</i>)	−0.0348	0.0145	−2.3949	0.0180	138
Eurasian skylark (<i>Alauda arvensis</i>)	−0.0684	0.0189	−3.6164	0.0004*	125
Meadow pipit (<i>Anthus pratensis</i>)	−0.0299	0.0184	−1.6273	0.1053	200
Yellowhammer (<i>Emberiza citrinella</i>)	−0.0385	0.0179	−2.1578	0.0367	44
Icterine warbler (<i>Hippolais icterina</i>)	−0.0705	0.0313	−2.2501	0.0285	57
Barn swallow (<i>Hirundo rustica</i>)	−0.2313	0.0544	−4.2540	0.0007*	17
Yellow wagtail (<i>Motacilla flava</i>)	−0.1255	0.0272	−4.6145	0.0000*	124
Tree sparrow (<i>Passer montanus</i>)	−0.1301	0.0815	−1.5971	0.1211	31
Willow warbler (<i>Phylloscopus trochilus</i>)	−0.0036	0.0094	−0.3827	0.7025	154
Stonechat (<i>Saxicola rubicola</i>)	−0.0279	0.0211	−1.3241	0.1891	85
Common starling (<i>Sturnus vulgaris</i>)	−0.1070	0.0315	−3.3991	0.0013*	57
Common whitethroat (<i>Sylvia communis</i>)	−0.0408	0.0125	−3.2751	0.0013*	179
Mistle thrush (<i>Turdus viscivorus</i>)	−0.1093	0.0277	−3.9480	0.0003*	44

Effect of imidacloprid concentration on annual intrinsic rate of increase in individual insectivorous bird species populations in the Netherlands.

*Species whose population is significantly affected by imidacloprid, after Bonferroni correction.

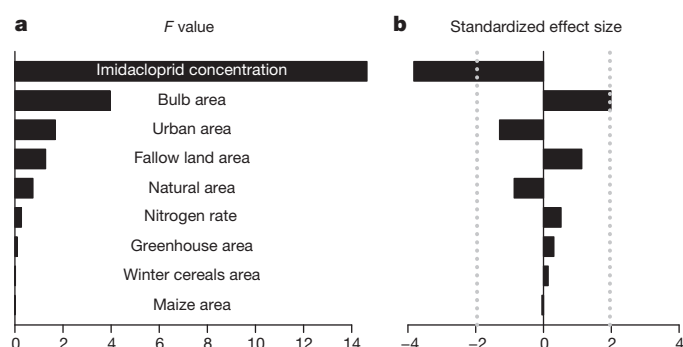


Figure 2 | Comparison of the effect of agricultural land-use changes and the effect of imidacloprid on bird population trends. **a**, The marginal variance ratio (F) of each effect was estimated from a mixed effects model with all species data pooled. **b**, The standardized effect size (t value) for each covariate from the mixed effects model. The vertical dotted lines represent significance thresholds at $\alpha = 0.05$ (two-sided test). The imidacloprid concentrations and the proportional changes in bulb production areas were the only variables that had significant effects (LMER: d.f. = 1,349, $t = -3.825$, $P = 0.0001$ for imidacloprid; and $t = 1.989$, $P = 0.0468$ for bulbs).

consumption of contaminated invertebrates and, for the six partly granivorous species involved, sublethal or lethal effects through the ingestion of coated seeds⁸. The relative effect sizes of these pathways urgently need to be investigated.

Farmland birds have experienced tremendous population declines in Europe in the past three decades, with agricultural intensification as the primary causal factor^{19–22}. Among aspects of intensification, pesticides are known to be a major threat to farmland birds^{15,23,24}. Neonicotinoids have recently replaced older intensively used insecticides such as carbamates, pyrethroids and organophosphates. After neonicotinoids were introduced to the Netherlands in the mid-1990s, their application was intensified, and the concentrations found in the environment frequently exceeded environmental standards, despite these concentrations being shown to have severe detrimental effects on several insect communities. Our results on the declines in bird populations suggest that neonicotinoids pose an even greater risk than has been anticipated. Cascading trophic effects deserve more attention in research on the ecosystem effects of this class of insecticides and must be taken into account in future legislation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 March; accepted 27 May 2014.

Published online 9 July 2014.

- Gill, R. J., Ramos-Rodriguez, O. & Raine, N. E. Combined pesticide exposure severely affects individual- and colony-level traits in bees. *Nature* **491**, 105–108 (2012).
- Henry, M. *et al.* A common pesticide decreases foraging success and survival in honey bees. *Science* **336**, 348–350 (2012).

- Whitehorn, P. R., O'Connor, S., Wackers, F. L. & Goulson, D. Neonicotinoid pesticide reduces bumble bee colony growth and queen production. *Science* **336**, 351–352 (2012).
- Van Dijk, T. C., van Staalduinen, M. A. & van der Sluijs, J. P. Macro-invertebrate decline in surface water polluted with imidacloprid. *PLoS ONE* **8**, e62374 (2013).
- Easton, A. H. & Goulson, D. The neonicotinoid insecticide imidacloprid repels pollinating flies and beetles at field-realistic concentrations. *PLoS ONE* **8**, e54819 (2013).
- Roessink, I., Merga, L. B., Zweers, H. J. & van den Brink, P. J. The neonicotinoid imidacloprid shows high chronic toxicity to mayfly nymphs. *Environ. Toxicol. Chem.* **32**, 1096–1100 (2013).
- Cramp, S. & Perrins, C. M. *The Birds of the Western Palearctic* (Oxford Univ. Press, 1994).
- Goulson, D. An overview of the environmental risks posed by neonicotinoid insecticides. *J. Appl. Ecol.* **50**, 977–987 (2013).
- Matsuda, K. *et al.* Neonicotinoids: insecticides acting on insect nicotinic acetylcholine receptors. *Trends Pharmacol. Sci.* **22**, 573–580 (2001).
- Tomizawa, M. & Casida, J. E. Neonicotinoid insecticide toxicology: mechanisms of selective action. *Annu. Rev. Pharmacol. Toxicol.* **45**, 247–268 (2005).
- Pollak, P. *Fine Chemicals: The Industry and the Business* (Wiley, 2011).
- Statistics Netherlands. *StatLine Databank* <http://statline.cbs.nl/statweb/> (2013).
- Institute of Environmental Sciences, Leiden University and Rijkswaterstaat-Water Services. *Dutch Pesticides Atlas* <http://www.bestrijdingsmiddelenatlas.nl> (2009).
- Newton, I. *Population Limitation in Birds* (Elsevier, 1998).
- Boatman, N. D. *et al.* Evidence for the indirect effects of pesticides on farmland birds. *Ibis* **146**, 131–143 (2004).
- Poulin, B., Lefebvre, G. & Paz, L. Red flag for green spray: adverse trophic effects of Bti on breeding birds. *J. Appl. Ecol.* **47**, 884–889 (2010).
- Van Turnhout, C. A. M., Foppen, R. P. B., Leuven, R. S. E. W., van Strien, A. & Siepel, H. Life-history and ecological correlates of population change in Dutch breeding birds. *Biol. Conserv.* **143**, 173–181 (2010).
- Van Turnhout, C. A. M., Foppen, R. P. B., Leuven, R. S. E. W., Siepel, H. & Esselink, H. Scale-dependent homogenization: changes in breeding bird diversity in the Netherlands over a 25-year period. *Biol. Conserv.* **134**, 505–516 (2007).
- Newton, I. The recent declines of farmland bird populations in Britain: an appraisal of causal factors and conservation actions. *Ibis* **146**, 579–600 (2004).
- Chamberlain, D. E. & Fuller, R. J. Local extinctions and changes in species richness of lowland farmland birds in England and Wales in relation to recent changes in agricultural land-use. *Agric. Ecosyst. Environ.* **78**, 1–17 (2000).
- Fuller, R. J. in *Ecology and Conservation of Lowland Farmland Birds* (eds Aebischer, N. J., Evans, A. D., Grice, P. V. & Vickery, J. A.) 5–16 (British Ornithologists' Union, 2000).
- Gregory, R. D. *et al.* Developing indicators for European birds. *Phil. Trans. R. Soc. Lond. B* **360**, 269–288 (2005).
- Geiger, F. *et al.* Persistent negative effects of pesticides on biodiversity and biological control potential on European farmland. *Basic Appl. Ecol.* **11**, 97–105 (2010).
- Mineau, P. & Whiteside, M. Pesticide acute toxicity is a better correlate of US grassland bird declines than agricultural intensification. *PLoS ONE* **8**, e57457 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. J. van Dijk, H. Sierdsema and D. Zoetebier for processing the bird data sets and J. P. van der Sluijs for sharing the database with imidacloprid concentration measurements. The Breeding Bird Monitoring Program is organised in close collaboration with Statistics Netherlands and provinces and is funded by the Dutch Ministry of EZ. We thank Sovon volunteers for their efforts in the field. The study was supported by NWO grants 840.11.001 and 841.11.007 and was the result of a collaboration within the Center for Avian Population Studies.

Author Contributions C.A.H. performed the statistical analysis. C.A.H., R.P.B.F., C.A.M.v.T., H.d.K. and E.J. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.A.H. (c.hallmann@science.ru.nl).

METHODS

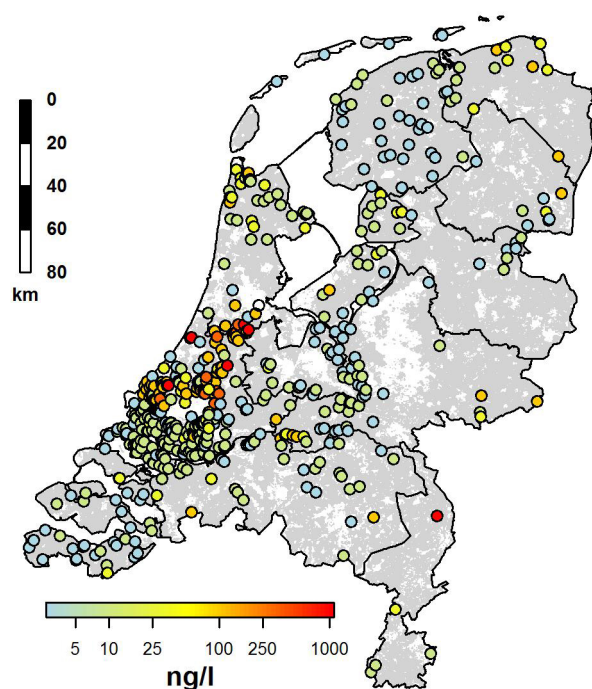
Data. We derived population trends for 15 insectivorous farmland passerine species (see Supplementary Data, Supplementary Methods and Extended Data Table 1 for the list of species) using long-term breeding bird data from the Dutch Common Breeding Bird Monitoring Scheme, a standardized^{25,26} monitoring scheme maintained and coordinated by Sovon, Dutch Centre for Field Ornithology, in collaboration with Statistics Netherlands¹⁷. The scheme has been running in the Netherlands since 1984. Data originating from these monitoring plots are generally considered to be adequately representative and reliable for population trend estimation^{17,18,25,27,28}. The monitoring plots are well scattered throughout the Netherlands and range in size between 10 ha and 1,000 ha (Extended Data Fig. 2).

We used previously described information on imidacloprid concentrations in Dutch surface water⁴. This data set was collected by the Dutch waterboard authorities as part of the regular monitoring of surface-water pesticide contamination¹³ (see Supplementary Data for details). Imidacloprid concentration measurements throughout the Netherlands are available (Extended Data Fig. 1); hence, this data set is considered an adequate representation of the actual water contamination levels in the Netherlands. The geographical locations of the two monitoring programs do not generally spatially coincide. To combine the data sets, we interpolated imidacloprid concentrations from water quality measurement locations to bird monitoring plots (see Supplementary Data).

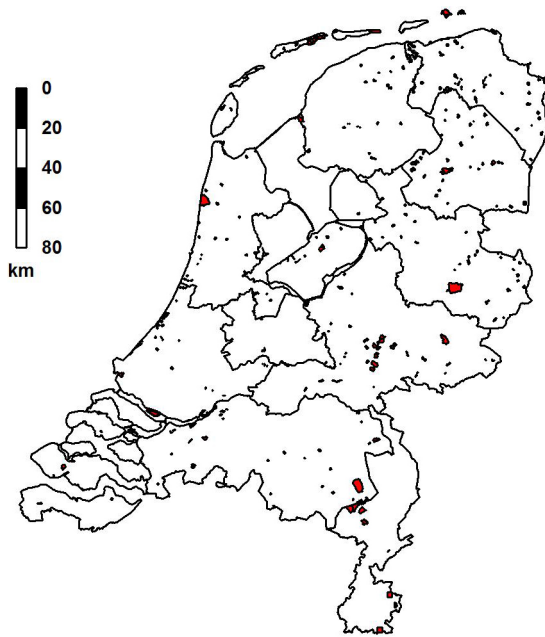
Statistical analysis. To assess the overall effects of expected concentrations on all species simultaneously, we used linear mixed effects models with species- and plot-specific population trends (intrinsic rates of increase or $\log[\lambda]$) as the response, \log [concentration of (interpolated) imidacloprid] as the fixed explanatory variable and species as a random factor. Additionally, we performed linear regressions of the population trends against the logarithm of the imidacloprid concentrations for each species separately using weighted least squares. The trends per plot were weighted by the mean species population size of the plot, to avoid the large influence of the demographic stochasticity of small populations. Population trends were calculated

as the slope of \log [territory counts] versus year of sampling (that is, a continuous trend) (see Supplementary Data). Regressions were performed using all monitoring plots located less than 5 km between the edge of a plot and an imidacloprid measurement location. This cut-off point of 5 km balanced the preferable proximity between bird and imidacloprid measurements with the amount of data retained in the analyses. However, regardless of how we varied the cut-off value between 1 and 25 km (that is, including between 7% and 99% of the bird monitoring plots, respectively), the effect size of imidacloprid on bird population trends remained strongly significantly negative (see Supplementary Methods and Extended Data Fig. 5). We examined potential confounding of the spatial imidacloprid concentrations with several different candidate explanatory variables that have been postulated as possible causes of farmland bird declines¹⁹ and that are relevant to the Netherlands¹⁷. We used eight variables¹² that are potentially confounded with the introduction of imidacloprid: namely, proportional change in the area of maize, proportional change in winter cereal cropping area, proportional change in flower bulb area, change in the amount of fertilizer application (nitrogen in kg ha^{-1}), proportional change in greenhouse area, proportional change in urban area, proportional change in natural habitat area and proportional change in fallow land area (Supplementary Data). We compared the significance of all explanatory variables using a multiple mixed effects model (with species intercept as a random effect) paired with F tests based on single term deletions of the full model (Fig. 2a). In addition, we compared standardized effect sizes (coefficient/s.e.m.) between explanatory variables based on single species multiple linear regression models (Fig. 2b and Supplementary Methods).

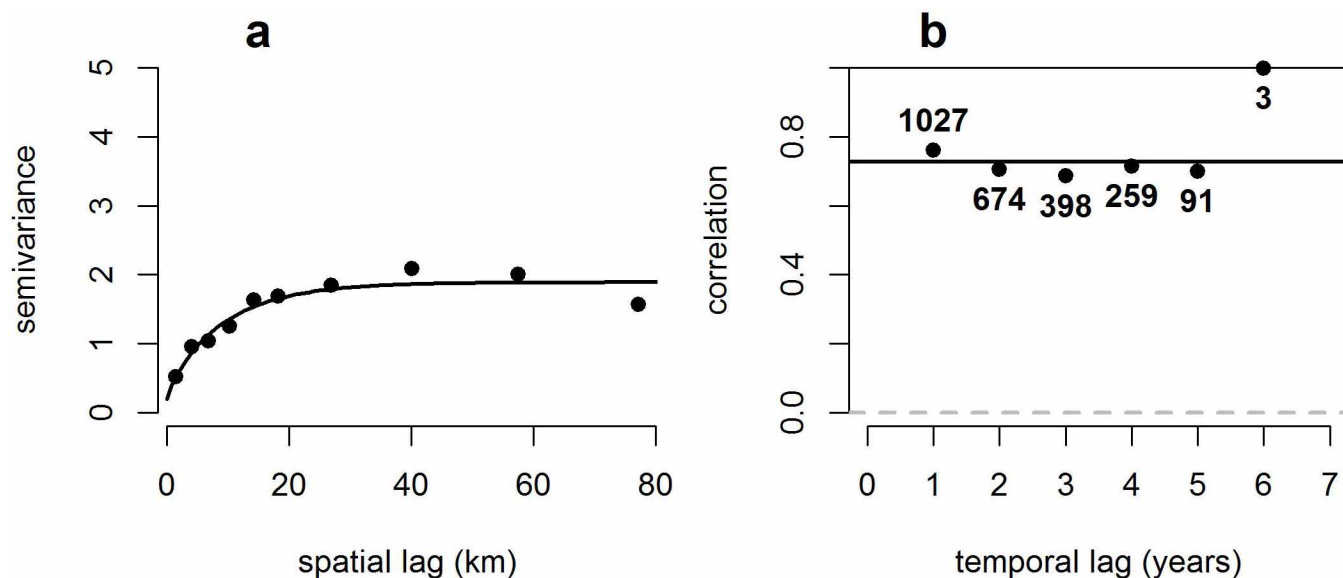
25. Bibby, C. J., Burgess, N. D. & Hill, D. A. *Bird Census Techniques* (Academic, 1992).
26. van Dijk, A. J. *Handleiding Broedvogel Monitoring Project* (Sovon Vogelonderzoek Nederland, 2004).
27. Devictor, V. *et al.* Differences in the climatic debts of birds and butterflies at a continental scale. *Nature Clim. Chang.* **2**, 121–124 (2012).
28. Kampichler, C., van Turnhout, C. A. M., Devictor, V. & van der Jeugd, H. P. Large-scale changes in community composition: determining land use and climate change signals. *PLoS ONE* **7**, e35272 (2012).



Extended Data Figure 1 | Distribution of the 555 imidacloprid measurement averages over the period 2003–2009, as used in the main analysis. The data are taken from refs 4 and 13.

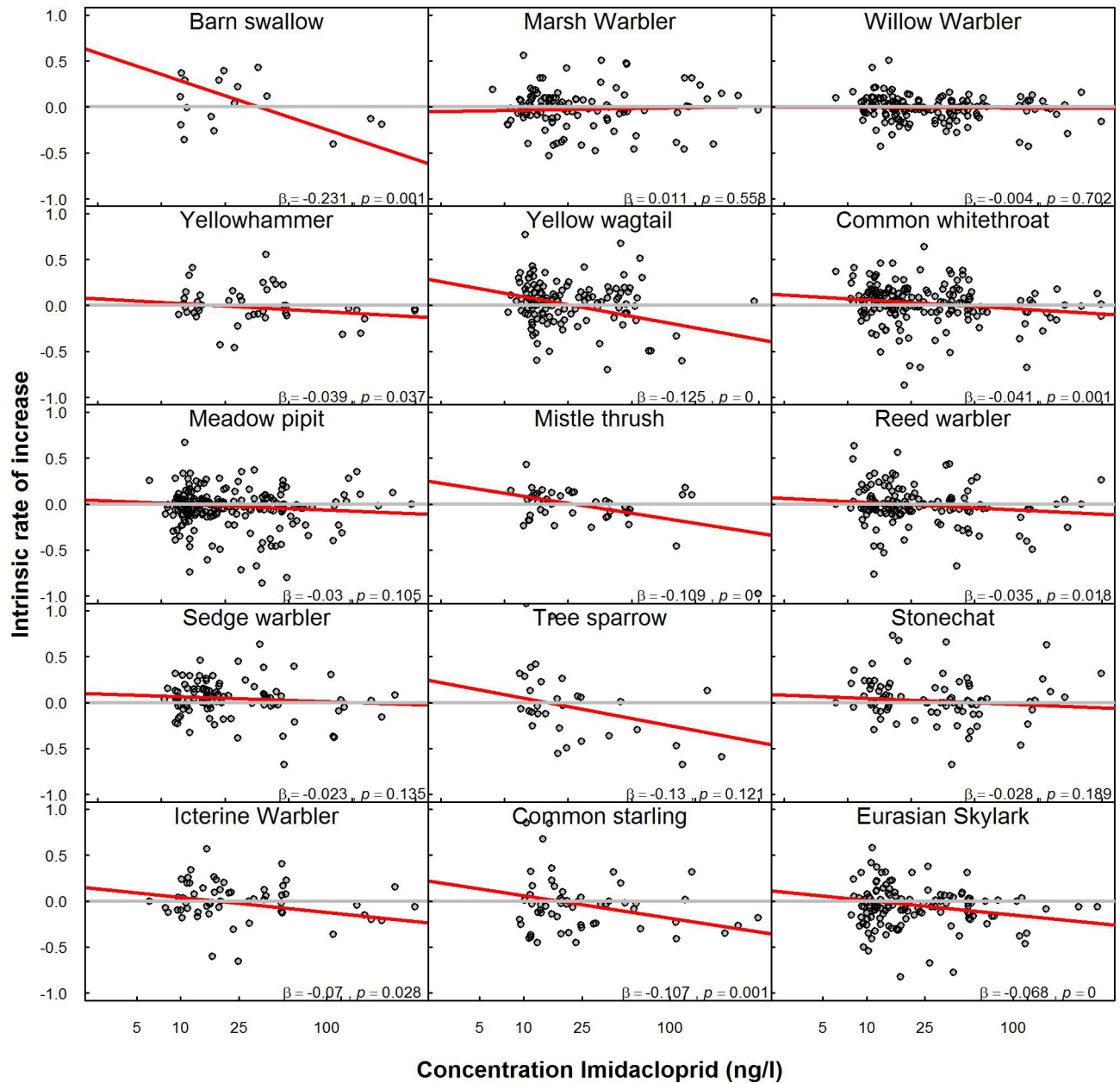


Extended Data Figure 2 | Distribution of the 354 bird monitoring plots in the Netherlands. The figure depicts the spatial distribution of bird monitoring plots from which local species-specific trends were calculated.

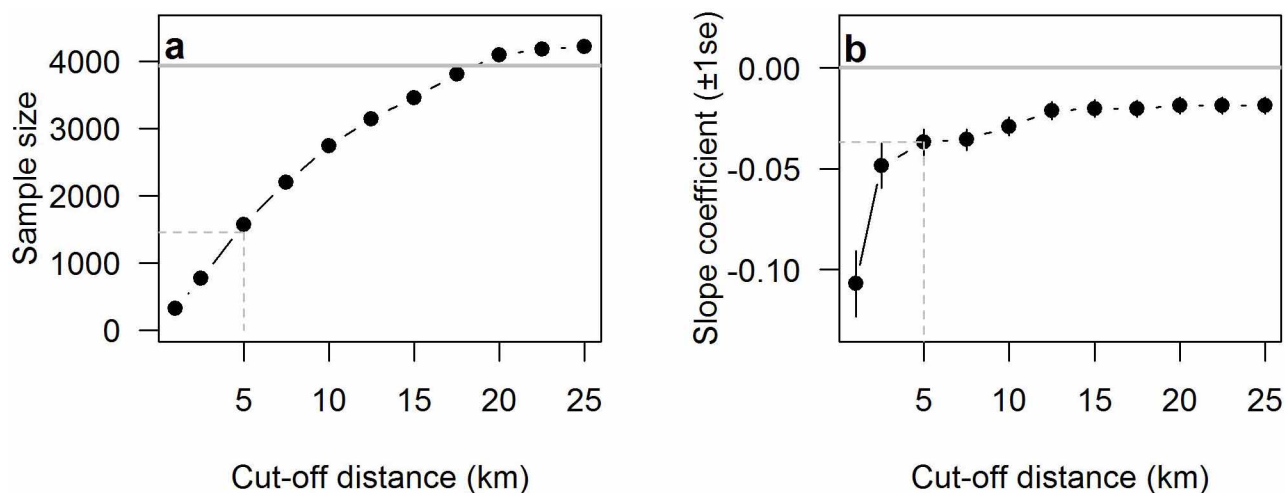


Extended Data Figure 3 | Spatial and serial (yearly) autocorrelation of imidacloprid measurements. **a**, Semivariance (dots) and Matern variogram model (fitted line) used in the interpolation of the concentrations (nugget = 0.1901, sill = 1.6989, range = 13.2 km). **b**, Serial correlation

(between years) of imidacloprid concentrations. Each value gives the number of pairs of measurements at each year lag that were used to calculate the coefficients. Serial correlations remain invariant with respect to temporal lag, indicating high temporal consistency in local imidacloprid concentrations.

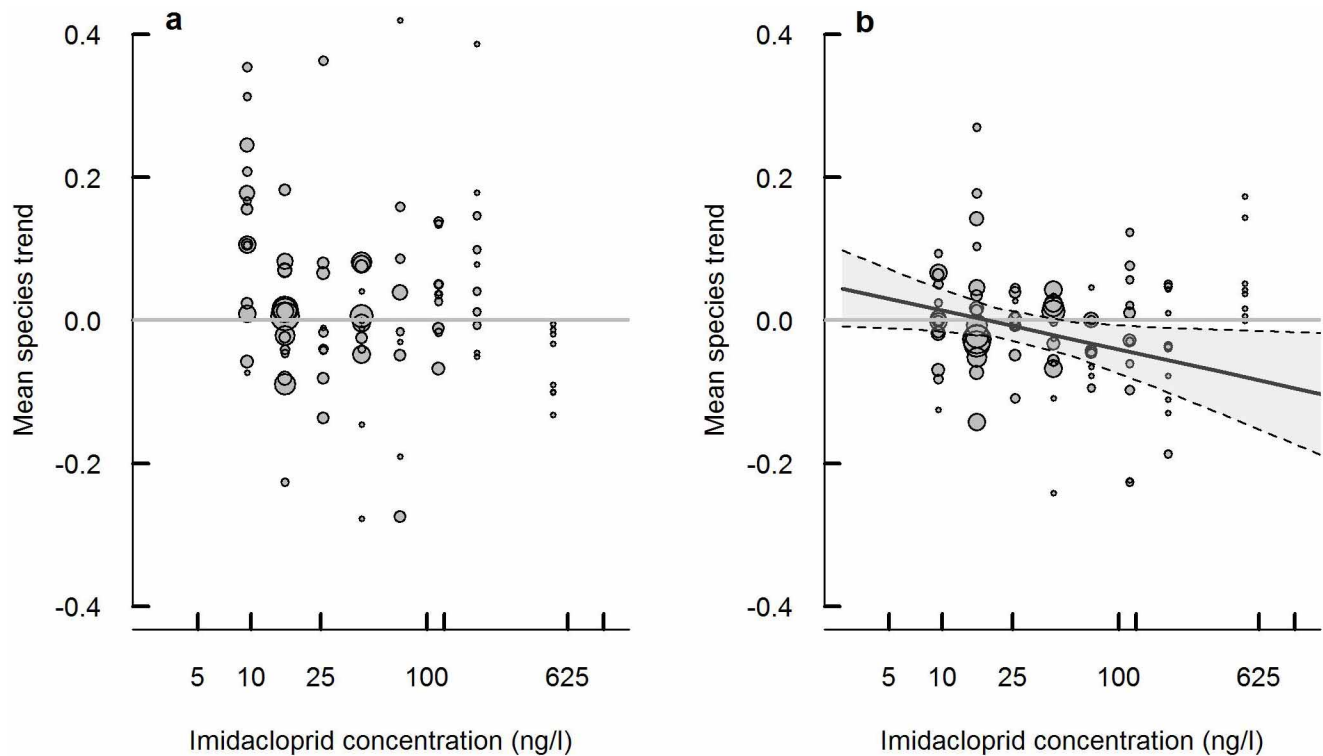


Extended Data Figure 4 | Population trends as a function of imidacloprid concentration per individual bird species. The red lines depict the weighted mean trend, also given as slope coefficients (β) and with corresponding P values.



Extended Data Figure 5 | Robustness check for the effect of the cut-off value for the distance between bird monitoring plots and water measurement locations (varied between 1 and 25 km). The larger the cut-off distance, the more species–plot annual rates of increase are retained in the analysis subset of

the total database of 3,947 records (a) but at the cost of increased noise in the response and a decrease in the effect of imidacloprid on the bird trends (b). However, in all cases, the effect of imidacloprid was significant and negative ($P < 0.0001$).



Extended Data Figure 6 | Bird species trends before and after imidacloprid introduction. Comparison of the relationship of bird species trends in the periods 1984–1995 (a) and 2003–2010 (b) with the imidacloprid concentrations in 2003–2009, based on all plots monitored in both time periods. Each point in the scatter plot represents the average intrinsic rate of increase of a species over all plots in the same concentration class. Binning into

classes was performed to reduce scatter noise and aid in visual interpretation. The actual analyses and the depicted significant regression line were based on raw data. The bird trends were significantly affected by the imidacloprid concentration in 2003–2010 ($t = -2.16$, d.f. = 365, $P = 0.031$) but were not significantly affected in the period before imidacloprid administration ($t = -1.43$, d.f. = 365, $P = 0.15$).

Extended Data Table 1 | Species information

Species	Foraging habitat	Migratory behaviour	Trend 1990-2005
Marsh warbler (<i>Acrocephalus palustris</i>)	reed	long-distance	stable
Sedge warbler (<i>Acrocephalus schoenobaenus</i>)	reed	long-distance	strong increase
Reed warbler (<i>Acrocephalus scirpaceus</i>)	reed	long-distance	stable
Eurasian skylark (<i>Alauda arvensis</i>)	farmland/grassland	short-distance	strong decline
Meadow pipit (<i>Anthus pratensis</i>)	grassland	short-distance	moderate decline
Yellowhammer (<i>Emberiza citrinella</i>)	farmland	resident	moderate increase
Icterine warbler (<i>Hippolais icterina</i>)	gardens/farms	long-distance	moderate decline
Barn swallow (<i>Hirundo rustica</i>)	farmland/grassland	long-distance	stable
Yellow wagtail (<i>Motacilla flava</i>)	farmland	long-distance	moderate decline
Tree sparrow (<i>Passer montanus</i>)	farmland	resident	stable
Willow warbler (<i>Phylloscopus trochilus</i>)	shrubs	long-distance	moderate decline
Stonechat (<i>Saxicola rubicola</i>)	shrubs	short-distance	strong increase
Common starling (<i>Sturnus vulgaris</i>)	grassland	short-distance	moderate decline
Common whitethroat (<i>Sylvia communis</i>)	shrubs	long-distance	moderate increase
Mistle thrush (<i>Turdus viscivorus</i>)	grassland	short-distance	moderate decline

Extended Data Table 2 | Multiple mixed effects regression of population trends (pooled over 15 species, $n = 1,926$)

	Coefficient(se)	<i>t</i> -value	<i>P</i> -value
Intercept	0.0932(0.0262)	3.5500	0.0004
Imidacloprid concentration	-0.0294(0.0077)	-3.8254	0.0001
Bulb area	0.0063(0.0032)	1.9895	0.0468
Urban area	-0.2970(0.2293)	-1.2954	0.1954
Fallow land area	1.2899(1.1428)	1.1287	0.2592
Natural area	-0.1878(0.2173)	-0.8646	0.3874
Nitrogen rates	$1.15(2.22) \times 10^{-5}$	0.5174	0.6050
Greenhouse area	0.0409(0.1340)	0.3050	0.7604
Winter cereals area	0.0543(0.3950)	0.1375	0.8906
Maize area	-0.0095(0.2062)	-0.0463	0.9631

Explanatory variables include log[imidacloprid concentration] (ng l^{-1}) and the area coverage change (difference in proportion of area, see Supplementary Data) of six land-use variables related to agricultural intensification and two variables potentially confounded with imidacloprid concentrations. For each explanatory variable, we present the slope coefficient along with the s.e.m., *t* and *P* values.

Genome sequencing identifies major causes of severe intellectual disability

Christian Gilissen^{1*}, Jayne Y. Hehir-Kwa^{1*}, Djie Tjwan Thung¹, Maartje van de Vorst¹, Bregje W. M. van Bon¹, Marjolien H. Willemsen¹, Michael Kwint¹, Irene M. Janssen¹, Alexander Hoischen¹, Annette Schenck¹, Richard Leach², Robert Klein², Rick Tearle², Tan Bo^{1,3}, Rolph Pfundt¹, Helger G. Yntema¹, Bert B. A. de Vries¹, Tjitske Kleefstra¹, Han G. Brunner^{1,4*}, Lisenka E. L. M. Vissers^{1*} & Joris A. Veltman^{1,4*}

Severe intellectual disability (ID) occurs in 0.5% of newborns and is thought to be largely genetic in origin^{1,2}. The extensive genetic heterogeneity of this disorder requires a genome-wide detection of all types of genetic variation. Microarray studies and, more recently, exome sequencing have demonstrated the importance of *de novo* copy number variations (CNVs) and single-nucleotide variations (SNVs) in ID, but the majority of cases remain undiagnosed^{3–6}. Here we applied whole-genome sequencing to 50 patients with severe ID and their unaffected parents. All patients included had not received a molecular diagnosis after extensive genetic prescreening, including microarray-based CNV studies and exome sequencing. Notwithstanding this prescreening, 84 *de novo* SNVs affecting the coding region were identified, which showed a statistically significant enrichment of loss-of-function mutations as well as an enrichment for genes previously implicated in ID-related disorders. In addition, we identified eight *de novo* CNVs, including single-exon and intra-exonic deletions, as well as interchromosomal duplications. These CNVs affected known ID genes more frequently than expected. On the basis of diagnostic interpretation of all *de novo* variants, a conclusive genetic diagnosis was reached in 20 patients. Together with one compound heterozygous CNV causing disease in a recessive mode, this results in a diagnostic yield of 42% in this extensively studied cohort, and 62% as a cumulative estimate in an unselected cohort. These results suggest that *de novo* SNVs and CNVs affecting the coding region are a major cause of severe ID. Genome sequencing can be applied as a single genetic test to reliably identify and characterize the comprehensive spectrum of genetic variation, providing a genetic diagnosis in the majority of patients with severe ID.

Whole-genome sequencing (WGS) is considered to be the most comprehensive genetic test so far⁷, but widespread application to patient diagnostics has been hampered by challenges in data analysis, the unknown diagnostic potential of the test, and relatively high costs. In this study, the genomes of 50 patients with severe ID and their unaffected parents were sequenced to an average genome-wide coverage of 80 fold (Supplementary Table 1)⁸. Before inclusion in the study, patients underwent an extensive clinical and genetic work-up, including targeted gene analysis, genomic microarray analysis and whole-exome sequencing (WES)⁶, but no molecular diagnosis could be established (Fig. 1).

On average, 98% of the genome was called for both alleles, giving rise to 4.4 million SNVs and 276 CNVs per genome (Supplementary Table 2). WGS identified an average of 22,186 coding SNVs per individual, encompassing more than 97% of variants identified previously by WES (Supplementary Tables 2, 3). We focused our analysis first on *de novo* SNVs and CNVs because of their importance in ID⁴. On average, 82 high-confidence potential *de novo* SNVs were called per genome (Supplementary Methods and Supplementary Table 4), which is in

concordance with previous studies^{9–11}. Systematic validation by Sanger sequencing of putative *de novo* variants in the protein-coding regions resulted in a total of 84 coding *de novo* mutations in 50 patients, giving rise to a protein-coding *de novo* substitution rate of 1.58 (Supplementary Methods and Supplementary Tables 5, 6, 7, 8). This rate exceeds all previously published substitution rates^{11–15} obtained using WES (Supplementary Table 9), as well as inferred substitution rates ($P = 3.58 \times 10^{-5}$) (ref. 14). In addition, this set of *de novo* mutations is significantly enriched for loss-of-function mutations ($P = 1.594 \times 10^{-5}$; Supplementary Methods).

Next, we investigated whether *de novo* mutations occurred in genes that have previously been identified in other patients with ID and/or overlapping phenotypes such as autism, schizophrenia or epilepsy^{12–19}. To this end, we compiled two sets of genes, one set containing 528 genes harbouring mutations in at least five patients with ID (referred to as ‘known ID genes’) and one list containing 628 genes harbouring mutations in at least one, but less than five patients (referred to as ‘candidate ID genes’) (Supplementary Methods). It has recently been shown that Mendelian disease genes are less tolerant to functional genetic variation than genes that do not cause any known disease²⁰. In line with this, both the set of known ID genes and the set of candidate ID genes indeed showed significantly less tolerance for functional variation ($P < 1.0 \times 10^{-6}$ for both sets; Extended Data Fig. 1 and Supplementary Methods). Subsequent analysis of our 84 *de novo* mutations at the gene level revealed significantly more mutations in known ID genes than expected (nine genes, $P = 0.04$; Supplementary Table 10). Mutations in these known ID genes included four insertion/deletion events, two non-sense mutations and three highly conserved missense mutations, thereby

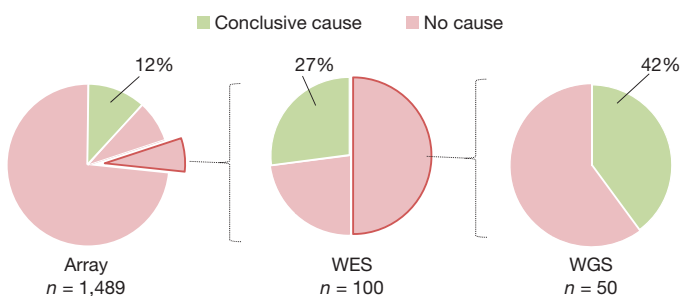


Figure 1 | Study design and diagnostic yield in patients with severe ID per technology. Diagnostic yield for patients with severe ID (IQ < 50), specified by technology: genomic microarrays, WES and WGS. Percentages indicate the number of patients in whom a conclusive cause was identified using the specified technique. Brackets indicate the group of patients in whom no genetic cause was identified and whose DNA was subsequently analysed using the next technology. WES data are updated with permission from ref. 6 (see Supplementary Methods).

¹Department of Human Genetics, Radboud Institute for Molecular Life Sciences and Donders Centre for Neuroscience, Radboud University Medical Center, Geert Grooteplein 10, 6525 GA Nijmegen, the Netherlands. ²Complete Genomics Inc. 2071 Stierlin Court, Mountain View, California 94043, USA. ³State Key Laboratory of Medical Genetics, Central South University. 110 Xiangya Road, Changsha, Hunan 410078, China. ⁴Department of Clinical Genetics, Maastricht University Medical Centre. Universiteitssingel 50, 6229 ER Maastricht, the Netherlands.

*These authors contributed equally to this work.

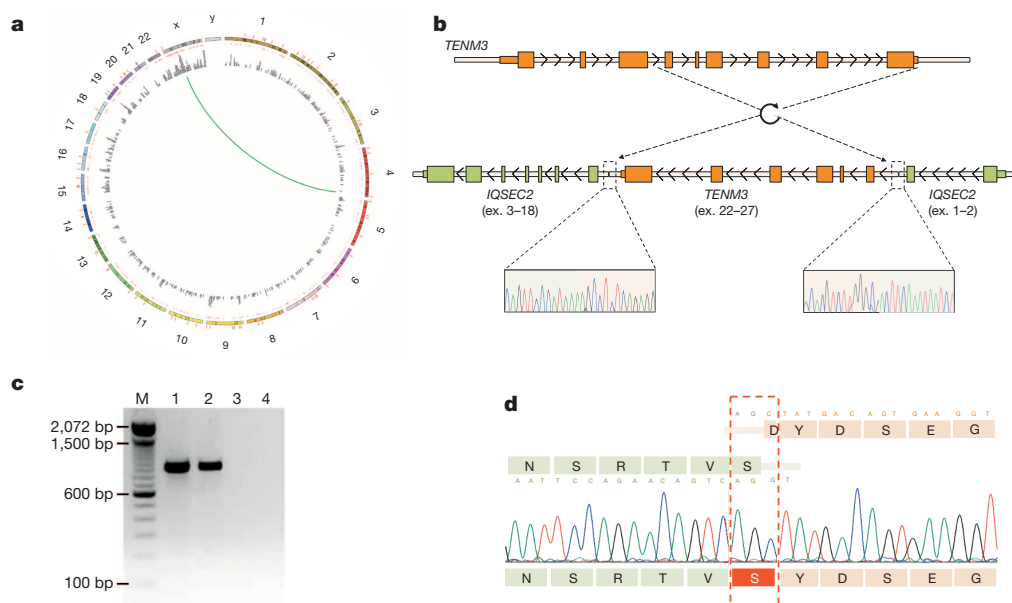


Figure 2 | Detected duplication of a chromosome 4 region into the X-chromosomal *IQSEC2* gene. **a–d**, Graphical representation of a *de novo* duplication–insertion event in patient 31. **a**, Circos plot with chromosome numbers and *de novo* mutations in the outer shell. Red bars represent genome-wide potential *de novo* SNVs, whereas blue lines represent potential *de novo* CNVs/structural variants. Inner shell represents the location of known ID genes (red marks) with the respective gene names. Green line illustrates a duplication event on chromosome 4, which is inserted into chromosome X. **b**, Details for inserted duplication event on chromosome X. The last six exons of

TENM3 are inserted in inverted orientation into intron 2 of *IQSEC2*, predicted to result in an in-frame *IQSEC2-TENM3* fusion gene. ex., exon. **c**, **d**, PCR (**c**) on and Sanger sequencing (**d**) of complementary DNA junction fragment in patient 31. Lanes in **c** represent the following: M, 100 bp marker; 1, cDNA of patient with cyclohexamide treatment; 2, cDNA of patient without cyclohexamide treatment; 3, control cDNA with cyclohexamide treatment; 4, control cDNA without cyclohexamide treatment. Our data verify the presence of a fusion gene in patient 31 that is suggested to escape nonsense-mediated decay.

also showing an enrichment for loss-of-function mutations ($P = 4.88 \times 10^{-6}$). Also, a significant enrichment for *de novo* mutations ($n = 10$) in candidate ID genes was identified ($P = 0.013$), including three loss-of-function mutations ($P = 0.02$) and three highly conserved missense mutations (Supplementary Table 11). These mutated known and candidate ID genes showed a diminished tolerance to functional variation ($P = 5.59 \times 10^{-6}$ and $P = 0.0042$, respectively), similar to what was observed for the entire set of known and candidate ID genes. These statistical analyses on SNVs together indicate that we not only identified significantly more *de novo* mutations in these 50 patients with severe ID, but also that they are more severe and occur more often in known or candidate ID genes.

In addition to the detection of *de novo* SNVs and small insertion/deletion events, a total of eight *de novo* structural variants, or CNVs, were identified and validated. These structural variants included five deletions, a tandem duplication, an interchromosomal duplication and one complex inversion/duplication/deletion event (Extended Data Table 1). All of these events had previously remained undetected by diagnostic microarray analysis. Three deletions were smaller than 10 kilobases (kb) in size, including two single-exon deletions and one intra-exonic deletion. Four of the *de novo* deletions encompassed a known ID gene and one a candidate ID gene, resulting in a significant enrichment for CNVs affecting known ID genes ($P = 0.015$). In addition, six *de novo* CNVs contained a gene in which exonic CNVs occur significantly more frequent in patients with ID ($n = 7,743$) compared to control individuals ($n = 4,056$) (Extended Data Table 1). Local realignment of sequence reads provided accurate single-nucleotide breakpoint information for six of the events, which was readily confirmed by breakpoint-spanning polymerase chain reactions (PCRs) (Extended Data Figs 2–5). Discordant reads not only provided the precise breakpoint sequences, but also positional information for duplicated sequences. In one case a partial duplication of *TENM3* on chromosome 4 was invertedly inserted into *IQSEC2* on the X chromosome. RNA studies confirmed the formation of a stable in-frame *IQSEC2-TENM3* gene fusion (Fig. 2), thereby suggesting that disruption

of *IQSEC2*, a known ID gene, may well contribute to the patient's phenotype. The contribution of such fusion genes to disease is well known in tumorigenesis, but has only recently been systematically investigated in neurodevelopmental disorders²¹.

Interestingly, three of ten *de novo* SNVs occurring in candidate ID genes seemed to be present in a mosaic state in the proband on the basis of the fraction of sequence reads containing the mutated allele. Sanger sequencing and amplicon-based deep sequencing confirmed the presence of mosaic mutations in these patients, at levels of 21% (*PIAS1*), 22% (*HIVEP2*) and 20% (*KANSL2*) (Extended Data Fig. 6), of which *KANSL2* is predicted to be deleterious owing to altered splicing. It is important that mosaic events like these can be detected by WGS as they are a known cause of genetic disease²². An additional advantage of genome sequencing over other approaches is that it may reveal pathogenic mutations in the non-coding part of the genome. In a systematic attempt to study the role of *de novo* non-coding mutations in ID, we selected all high-confidence candidate *de novo* mutations located either within the promoter regions, introns or untranslated regions of all known ID genes and validated 43 mutations (Supplementary Tables 12, 13). Annotation of these mutations using several ENCODE resources²³, including chromatin state segments of nine human cell types and transcription-factor-binding sites, did not reveal potential pathogenic non-coding mutations (Supplementary Methods). However, our understanding of non-coding variation is still limited and extensive functional follow-up will be required to determine its role in disease²³.

In addition to the statistical analysis of our data, we also assessed the impact of our genome sequencing study in a clinical diagnostic setting, in which variant interpretation is combined with an evaluation of patients' phenotypes to make a diagnostic decision on a per patient basis. Therefore, all *de novo* coding mutations (SNVs and CNVs) were evaluated for pathogenicity on the basis of established diagnostic criteria (Supplementary Methods and Extended Data Table 1)^{24–27}. For patients with *de novo* mutations in a known or candidate ID gene this clinical diagnostic assessment also included a comparison of the phenotype observed

Table 1 | Diagnostic yield by WGS for a pre-screened cohort of 50 ID trios

Genetic cause	Number of patients
Total positive diagnosis	21
Dominant <i>de novo</i>	20
Autosomal SNV	11
Autosomal CNV	5
X-linked SNV	2
X-linked CNV	2
Recessive	1
Homozygous	0
Compound heterozygous	1
X-linked	0
Candidate ID genes	8
No diagnosis	21

in our patient with those reported in the literature. Conclusive diagnoses were reached for fourteen patients with *de novo* mutations affecting a known ID gene (nine SNVs and five CNVs), as well as for six patients with *de novo* mutations affecting a candidate ID gene (four SNVs and two CNVs) (Extended Data Tables 1, 2 and Supplementary Tables 8, 14).

Although family history for ID was negative for all patients included in this study, we evaluated the presence of recessively inherited causes of disease due to mutations in known ID genes (Supplementary Table 10). We did not find X-linked maternally inherited variants in male patients consistent with the patient's phenotype, nor did we identify relevant homozygous or compound heterozygous SNVs on the autosomes. We did, however, identify a single proband carrying compound heterozygous deletions affecting the *VPS13B* gene, one of the known ID genes. Subsequent breakpoint sequencing confirmed that the 122 kb deletion, affecting exons 12–18, was paternally inherited whereas the 1.7 kb deletion of the last exon was maternally inherited (Extended Data Fig. 7). Notably, Cohen syndrome²⁸ was part of the differential diagnosis of this patient but no causative SNVs or CNVs were previously detected in this gene by direct Sanger sequencing or microarray analysis.

Taken together, a conclusive diagnosis was made in 21 of 50 patients with severe ID in this well-studied cohort (42%; Table 1 and Supplementary Table 15). The experimental set-up of our study allowed us to estimate the diagnostic yield of WGS in an unbiased cohort of such patients (Fig. 1). On the basis of established diagnostic rates for genomic

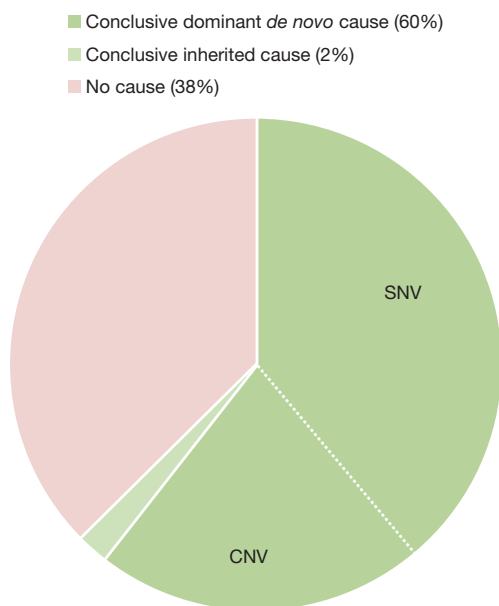


Figure 3 | Pie chart showing role of *de novo* mutations in severe ID. Contribution of genetic causes to severe ID on the basis of the cumulative estimates provided per technology. Our data indicate that *de novo* mutations are a major cause of severe ID. Note, small variants include SNVs and insertion/deletion events whereas large variants include structural variants and CNVs (>500 bp).

microarrays (12%) and WES (27%) in patients from the same large cohort^{6,26}, the cumulative estimate for WGS to reach a conclusive genetic diagnosis is 62%, of which 60% by *de novo* events (39% SNVs, 21% CNVs) and 2% by recessive inheritance (Fig. 3 and Supplementary Methods). The role of *de novo* somatic mutations and *de novo* mutations outside the coding regions remains to be fully explored.

METHODS SUMMARY

Patients were selected to have severe ID (IQ < 50) and negative results on diagnostic genomic microarrays and exome sequencing⁶ (Fig. 1). WGS was performed by Complete Genomics as previously described^{8,29}. *De novo* SNVs were identified using Complete Genomics' cgatools 'calldiff' program. CNVs and structural variants were reported by Complete Genomics on the basis of read-depth deviations and discordant read pairs, respectively. *De novo* CNVs and structural variants were then identified by excluding variants with minimal evidence or overlapping with CNVs and structural variants identified in the parents or control data sets. All variants were annotated using an in-house analysis pipeline and subsequently prioritized for validation based on their confidence level (low/medium/high) and location in the genome (coding/non-coding). High-confidence candidate *de novo* mutations in non-coding variants in known ID genes were prioritized on the basis of evolutionary conservation and overlap with ENCODE chromatin state segments and transcription-factor-binding sites²³. Statistical overrepresentation of mutations in known and candidate ID gene lists was calculated using Fisher's exact test based on RefSeq genes. Enrichments for loss-of-function CNV events were calculated using the exact Poisson test. To clinically interpret (*de novo*) mutations, each variant (both CNV and SNV) was assessed for mutation impact as well as functional relevance to ID according to diagnostic protocols for variant interpretation^{6,24–27}. The diagnostic yield of WGS in an unbiased cohort was calculated based on cumulative estimates of diagnostic yield per technology (genomic microarray, WES and WGS).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 March 2014; accepted 17 April 2014.

Published online 4 June 2014.

- Ropers, H. H. Genetics of early onset cognitive impairment. *Annu. Rev. Genomics Hum. Genet.* **11**, 161–187 (2010).
- Mefford, H. C., Batshaw, M. L. & Hoffman, E. P. Genomics, intellectual disability, and autism. *N. Engl. J. Med.* **366**, 733–743 (2012).
- de Vries, B. B. *et al.* Diagnostic genome profiling in mental retardation. *Am. J. Hum. Genet.* **77**, 606–616 (2005).
- Vissers, L. E. *et al.* A *de novo* paradigm for mental retardation. *Nature Genet.* **42**, 1109–1112 (2010).
- Rauch, A. *et al.* Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* **380**, 1674–1682 (2012).
- de Ligt, J. *et al.* Diagnostic exome sequencing in persons with severe intellectual disability. *N. Engl. J. Med.* **367**, 1921–1929 (2012).
- Lupski, J. R. *et al.* Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N. Engl. J. Med.* **362**, 1181–1191 (2010).
- Drmanac, R. *et al.* Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* **327**, 78–81 (2010).
- Michaelson, J. J. *et al.* Whole-genome sequencing in autism identifies hot spots for *de novo* germline mutation. *Cell* **151**, 1431–1442 (2012).
- Kong, A. *et al.* Rate of *de novo* mutations and the importance of father's age to disease risk. *Nature* **488**, 471–475 (2012).
- Jiang, Y. H. *et al.* Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. *Am. J. Hum. Genet.* **93**, 249–263 (2013).
- O'Roak, B. J. *et al.* Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations. *Nature* **485**, 246–250 (2012).
- Iossifov, I. *et al.* *De novo* gene disruptions in children on the autistic spectrum. *Neuron* **74**, 285–299 (2012).
- Neale, B. M. *et al.* Patterns and rates of exonic *de novo* mutations in autism spectrum disorders. *Nature* **485**, 242–245 (2012).
- Sanders, S. J. *et al.* *De novo* mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–241 (2012).
- Epi4K Consortium & Epilepsy Phenome/Genome Project. *De novo* mutations in epileptic encephalopathies. *Nature* **501**, 217–221 (2013).
- Gulsuner, S. *et al.* Spatial and temporal mapping of *de novo* mutations in schizophrenia to a fetal prefrontal cortical network. *Cell* **154**, 518–529 (2013).
- Xu, B. *et al.* Exome sequencing supports a *de novo* mutational paradigm for schizophrenia. *Nature Genet.* **43**, 864–868 (2011).
- Girard, S. L. *et al.* Increased exonic *de novo* mutation rate in individuals with schizophrenia. *Nature Genet.* **43**, 860–863 (2011).
- Petrovski, S., Wang, Q., Heinzen, E. L., Allen, A. S. & Goldstein, D. B. Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet.* **9**, e1003709 (2013).

21. Rippey, C. *et al.* Formation of chimeric genes by copy-number variation as a mutational mechanism in schizophrenia. *Am. J. Hum. Genet.* **93**, 697–710 (2013).
22. Biesecker, L. G. & Spinner, N. B. A genomic view of mosaicism and human disease. *Nature Rev. Genet.* **14**, 307–320 (2013).
23. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
24. Bell, J. B. D., Sistermans, E. & Ramsden, S. C. *Practice guidelines for the Interpretation and Reporting of Unclassified Variants (UVs) in Clinical Molecular Genetics* (The UK Clinical Molecular Genetics Society and the Dutch Society of Clinical Genetic Laboratory Specialists, 2007).
25. Berg, J. S., Khoury, M. J. & Evans, J. P. Deploying whole genome sequencing in clinical practice and public health: meeting the challenge one bin at a time. *Genet. Med.* **13**, 499–504 (2011).
26. Vulto-van Silfhout, A. T. *et al.* Clinical significance of *de novo* and inherited copy-number variation. *Hum. Mutat.* **34**, 1679–1687 (2013).
27. Hehir-Kwa, J. Y., Pfundt, R., Veltman, J. A. & de Leeuw, N. Pathogenic or not? Assessing the clinical relevance of copy number variants. *Clin. Genet.* **84**, 415–421 (2013).
28. Kolehmainen, J. *et al.* Cohen syndrome is caused by mutations in a novel gene, *COH1*, encoding a transmembrane protein with a presumed role in vesicle-mediated sorting and intracellular protein transport. *Am. J. Hum. Genet.* **72**, 1359–1369 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank R. Drmanac, K. Albers, J. Goeman, D. Lugtenberg and P. N. Robinson for useful discussions, and M. Steehouwer, P. de Vries and W. Nillesen for technical support. This work was in part financially supported by grants from the Netherlands Organization for Scientific Research (912-12-109 to J.A.V., A.S. and B.B.A.d.V., 916-14-043 to C.G., 916-12-095 to A.H., 907-00-365 to T.K. and SH-271-13 to C.G. and J.A.V.) and the European Research Council (ERC Starting grant DENOVO 281964 to J.A.V.).

Author Contributions Laboratory work: M.K., I.M.J., T.B., A.H., L.E.L.M.V. Clinical investigation: B.W.M.v.B., M.H.W., B.B.A.d.V., T.K., H.G.B. Data analysis: C.G., J.Y.H.-K., D.T.T., M.v.d.V., R.T. Generation of ID gene list: C.G., A.S., R.P., H.G.Y., T.K., L.E.L.M.V. Data interpretation: L.E.L.M.V., R.P., H.G.Y. Study design: J.A.V., H.G.B., R.L., R.K. Supervision of the study: H.G.B., L.E.L.M.V., J.A.V. Manuscript writing: C.G., J.Y.H.-K., H.G.B., L.E.L.M.V., J.A.V.

Author Information Data included in this manuscript have been deposited at the European Genome-phenome Archive (<https://www.ebi.ac.uk/ega/home>) under accession number EGAS00001000769. Reprints and permissions information is available at www.nature.com/reprints. Readers are welcome to comment on the online version of the paper. The authors declare competing financial interests: details are available in the online version of the paper. Correspondence and requests for materials should be addressed to J.A.V. (joris.veltman@radboudumc.nl).

METHODS

Patient selection. Patients were selected to have severe ID ($IQ < 50$) and negative results on diagnostic genomic microarrays and exome sequencing⁶ (Fig. 1 and Supplementary Methods).

Whole genome sequencing. WGS was performed by Complete Genomics as previously described⁸. Sequence reads were mapped to the reference genome (GRCh37) and variants were called by local *de novo* assembly according to the methods previously described²⁹.

Identification of *de novo* small variants. *De novo* SNVs were identified using Complete Genomics' cgatools 'calldiff' program. On the basis of the rank order of the two confidence scores of a *de novo* mutation, we binned the variants in three groups: low confidence (at least one score < 0), medium confidence (both scores ≥ 0 but at least one < 5) and high confidence (both scores ≥ 5) (Supplementary Methods).

Identification of X-linked, recessive and compound heterozygous SNVs. Maternally inherited X-linked variants (in male patients), homozygous variants and compound heterozygous variant pairs were identified using the Complete Genomics' cgatools 'listvariants' and 'testvariants' programs to select variants according to their respective segregation. Compound heterozygous variants affecting the same gene were identified using RefSeq gene annotation (Supplementary Methods).

Identification of *de novo* CNVs and structural variants. CNVs were reported by Complete Genomics on the basis of read-depth deviations across 2 kb windows. Structural variants were reported by Complete Genomics based on discordant read pairs. *De novo* CNVs/structural variants were then identified by excluding variants with minimal evidence or overlapping with CNVs/structural variants identified in the parents or control data sets (Supplementary Methods).

Generation of lists for known and candidate ID genes. To prioritize and for subsequent interpretation of *de novo* variants for each patient individually, two gene lists were generated, one containing known ID genes (defined by five or more patients with ID having a mutation in the respective gene) and one containing candidate ID genes (defined by at least one but less than five patients with ID (or a related phenotype) showing a mutation in the respective gene) (Supplementary Methods).

Prioritization of clinically relevant SNVs and CNVs or structural variants. All SNVs were annotated using an in-house analysis pipeline. Variants were prioritized

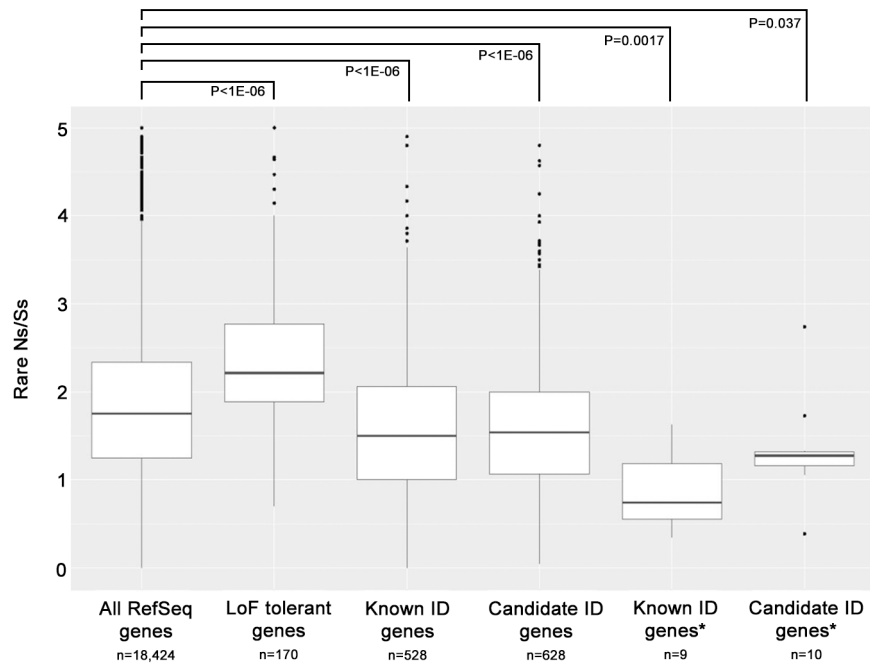
for validation in two distinct ways: (1) medium and high-confidence *de novo* SNVs and *de novo* CNVs/structural variants affecting coding regions and/or canonical splice sites; and (2) all potential *de novo* variants within known ID genes, irrespective of confidence level. Interpretation of coding *de novo* variants was performed as described previously⁶. High-confidence candidate *de novo* mutations in non-coding variants were prioritized on the basis of evolutionary conservation and overlap with ENCODE chromatin state segments and transcription-factor-binding sites (Supplementary Methods)²³.

Clinical interpretation of mutations. To clinically interpret (*de novo*) mutations, each *de novo* mutation (both CNV and SNV) was assessed for mutation impact as well as functional relevance to ID according to diagnostic protocols for variant interpretation^{24–27} that are used in our accredited diagnostic laboratory for genetic analysis (accredited to the 'CCKL Code of Practice', which is based on EN/ISO 15189 (2003), registration numbers R114/R115, accreditation numbers 095/103) (Supplementary Methods).

Statistical analysis. Overrepresentation of mutations in gene lists was calculated using Fisher's exact test based on the total coding size of all RefSeq genes and coding size of the genes from the respective gene list. Overrepresentation of loss-of-function mutations was calculated using Fisher's exact test based on published control cohorts. Enrichments for loss-of-function CNV events were calculated using the exact Poisson test. Enrichment for known ID genes was calculated using Fisher's exact test and odds ratios were calculated to compare the frequency of exonic CNVs in ID and control cohorts, respectively (Supplementary Methods).

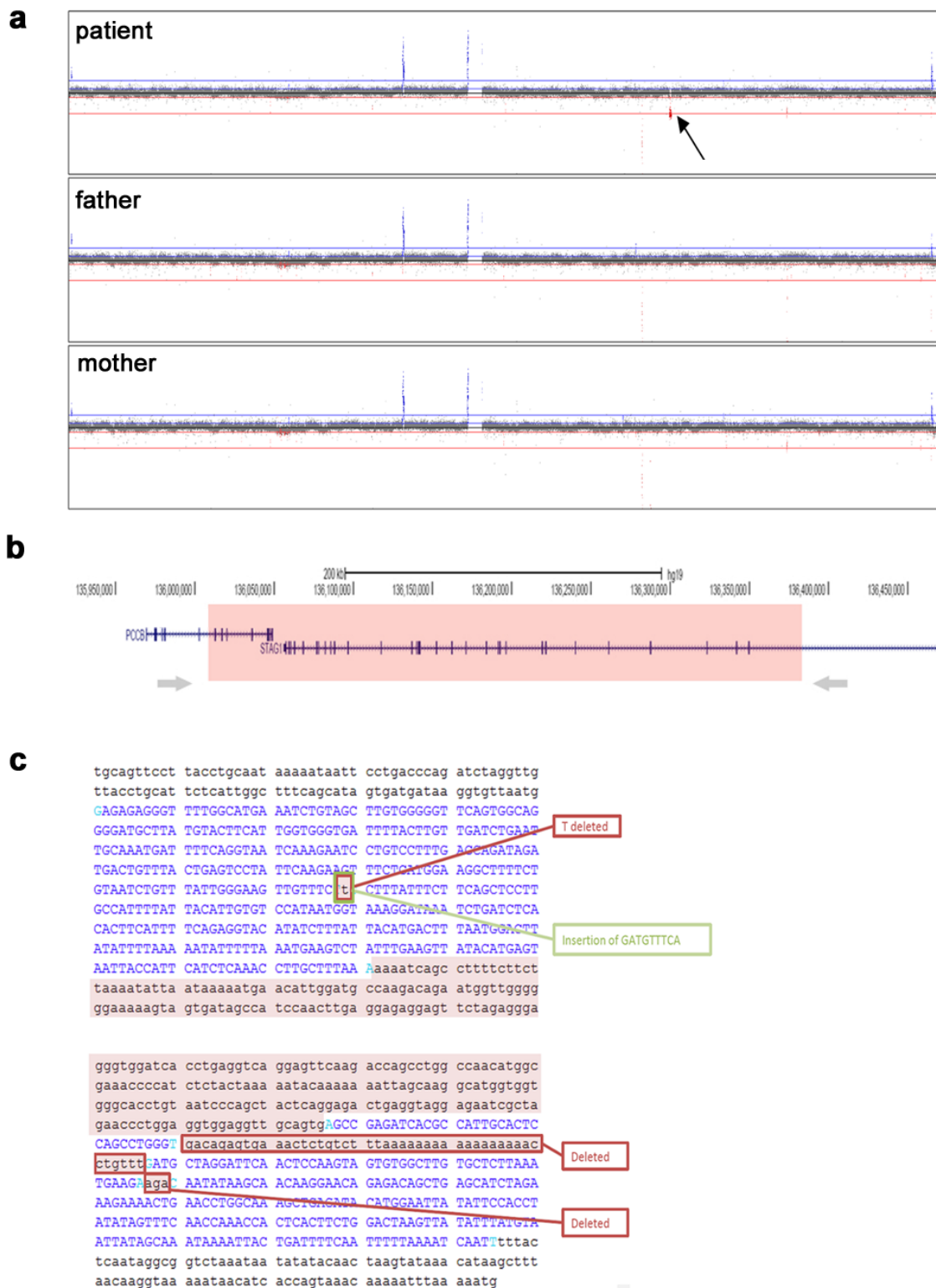
Calculation of diagnostic yield. Our in-house phenotypic database contains 1,489 patients with severe ID who have all had a diagnostic genomic microarray in the time period 2003–2013. In 173 (11.6%) of these patients, a *de novo* CNV was identified as a cause of ID. Subsequently, 100 array-negative patients were subjected to WES, which resulted in a *de novo* cause for ID in 27% of patients. Of all WES-negative patients, 50 were selected for this WGS study, in which 42% obtained a conclusive genetic cause. Cumulative estimates were subsequently determined using the diagnostic yield per technology (Supplementary Methods).

29. Carnevali, P. *et al.* Computational techniques for human genome resequencing using mated gapped reads. *J. Comput. Biol.* **19**, 279–292 (2012).
30. MacArthur, D. G. *et al.* A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**, 823–828 (2012).



Extended Data Figure 1 | Boxplots of rare missense burden in different gene sets. Boxplots showing the difference in tolerance for rare missense variation in the general population. The vertical axis shows the distribution for each gene set of the number of rare (<1% in NHLBI Exome Sequencing Project) missense variants divided by the number of rare synonymous variants. From left to right the following gene sets are depicted: all 18,424 RefSeq genes,

170 loss-of-function tolerant genes from ref. 30, all 528 known ID genes (Supplementary Table 10), all 628 candidate ID genes (Supplementary Table 11), 9 known ID genes in which *de novo* mutations were identified in this study (Supplementary Table 8), and 10 candidate ID genes in which *de novo* mutations were identified in this study (Supplementary Table 8).

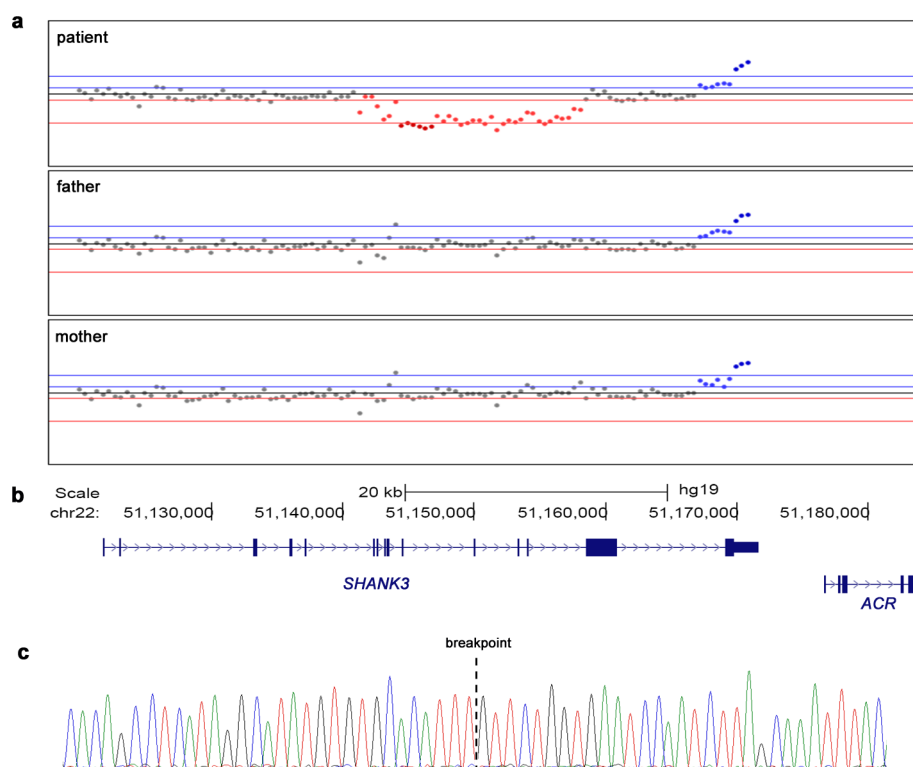


Extended Data Figure 2 | Structural variant involving *STAG1* (patient 40).

a–c, CNV identified using WGS in patient 40, including the *STAG1* gene.

a, Chromosome 3 profile (log₂ test over reference (T/R) ratios) based on read-depth information for patient, father and mother. Black arrow points

towards the *de novo* event in patient 40. **b**, Genic contents of deletion. Grey arrows show primers used to amplify the junction fragment. **c**, Details on the proximal and distal breakpoints, showing the ‘fragmented’ sequence at both ends. Breakpoints are provided in Extended Data Table 1.

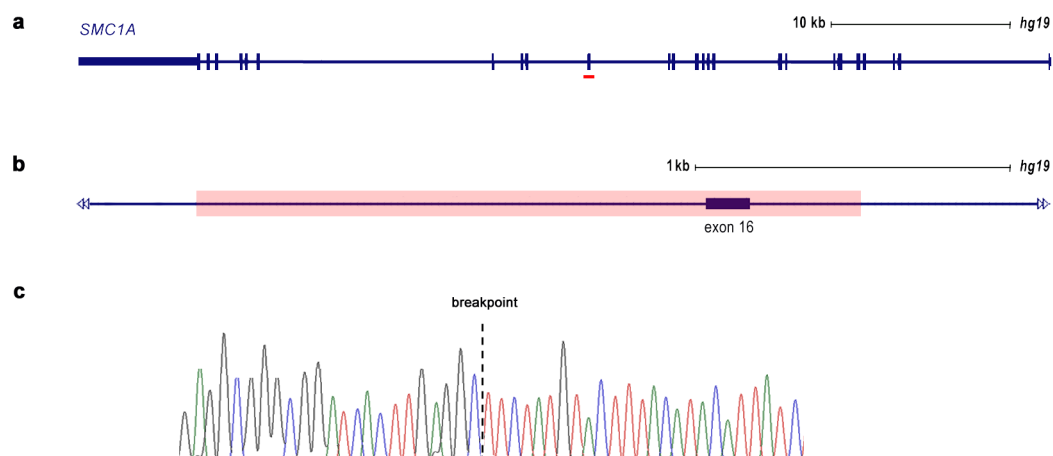


Extended Data Figure 3 | Structural variant involving *SHANK3* (patient 5).

a–c, CNV identified using WGS in patient 5, including the *SHANK3* gene.

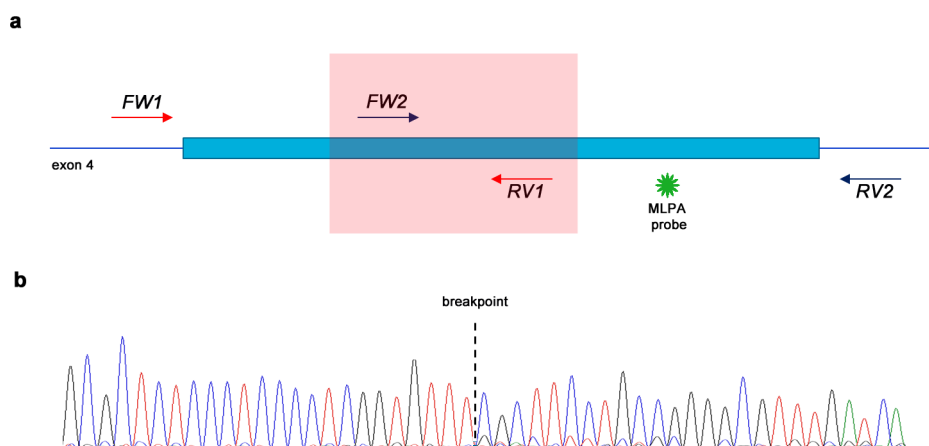
a, Detail of chromosome 22 profile (\log_2 T/R ratios) based on read-depth information for patient, father and mother. Red dots in top panel show ratios indicating the *de novo* deletion in patient 5. **b**, Genic content of the deletion.

c, Sanger validation for the junction fragment. Dotted vertical line indicates the breakpoint with sequence on the left side originating from sequence proximal to *SHANK3* and on the right side sequence that originates from sequence distal to *ACR*. Breakpoints are provided in Extended Data Table 1.



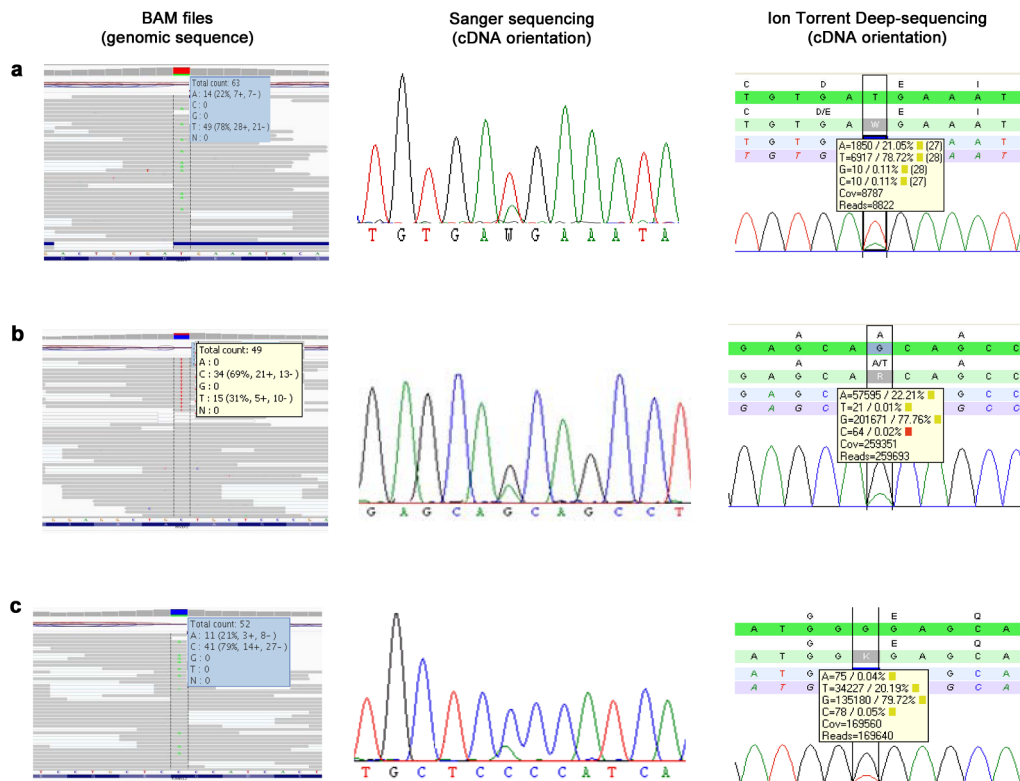
Extended Data Figure 4 | Single-exon deletion involving *SMC1A* (patient 48). **a**, Schematic depiction of the deletion identified in patient 48 involving a single exon of *SMC1A*. Pink horizontal bar highlights the exon that was deleted in the patient. **b**, Details at the genomic level of the deletion including

exon 16, with Sanger sequence validation of the breakpoints. Junction is indicated by a black vertical dotted line. Breakpoints are provided in Extended Data Table 1.



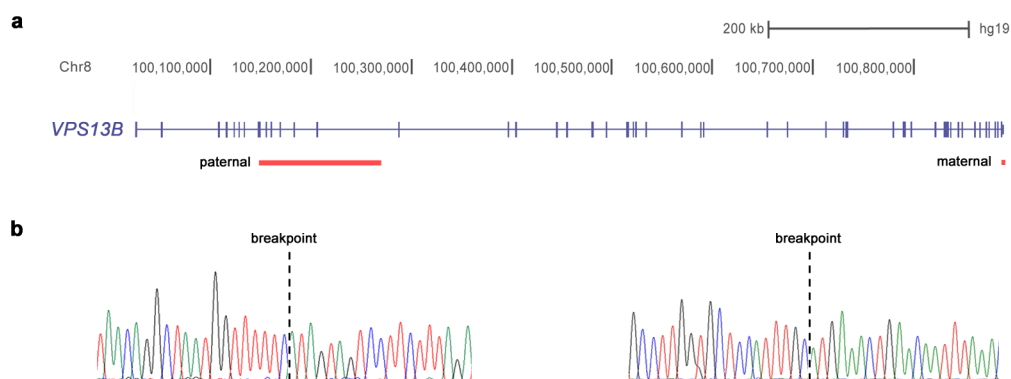
Extended Data Figure 5 | Intra-exonic deletion involving *MECP2* (patient 18). **a**, Schematic depiction of the deletion identified in patient 18, which is located within exon 4 of *MECP2*. Initial Sanger sequencing in a diagnostic setting could not validate the deletion as the primers used to amplify exon 4 removed the primer-binding sites (FW2 and RV1 respectively). Multiplex ligation probe amplification (MLPA) analysis for CNV detection showed

normal results as the MLPA primer-binding sites were located just outside of the deleted region. **b**, Combining primers FW1 and RV2 amplified the junction fragment, clearly showing the deletion within exon 4. Of note, the background underneath the Sanger sequence is derived from the wild-type allele. Breakpoints are provided in Extended Data Table 1.



Extended Data Figure 6 | Confirmation of mosaic mutations in *PIAS1*, *HIVEP2* and *KANSL2*. a–c, Approaches used to confirm the presence of mosaic mutations in *PIAS1* (a), *HIVEP2* (b) and *KANSL2* (c). Images and read-depth information showing the base counts in the BAM files (left) indicated that the variants/wild-type allele were not in a 50%/50% distribution. Sanger sequencing (middle) then confirmed the variant to be present in the patient,

and absent in the parents (data from parents not shown), again indicating that the mutation allele is underrepresented. Guided by these two observations, amplicon-based deep sequencing using Ion Torrent subsequently confirmed the mosaic state of the mutations (right). On the basis of deep sequencing, percentages of mosaicism for *PIAS1*, *HIVEP2* and *KANSL2* were estimated at 21%, 22% and 20%, respectively.



Extended Data Figure 7 | Compound heterozygous structural variation affecting *VPS13B* (patient 12). **a, b,** CNVs of *VPS13B* identified using WGS in patient 12. **a,** Schematic representation of *VPS13B*, with vertical bars indicating coding exons. In patient 12 two deletions were identified, one ~122 kb in size which was inherited from his father, and another ~2 kb in size, which was

inherited from his mother and consisted only of a single exon. **b,** Both CNV junction fragments were subsequently validated using Sanger sequencing. Left, junction fragment from the paternally inherited deletion. Right, junction fragment from the maternally inherited deletion. Breakpoints are provided in Extended Data Table 1.

Extended Data Table 1 | Large variants of potential clinical relevance identified using WGS and probability of exonic CNVs occurring in affected and control individuals for these loci

Trio	Type*	Genomic characterization	Size (kb)	CN	Origin	Genes affected	Affected (n=7,743)	Controls (n=4,056)	OR	CI	P-value#
5	CNV	chr22(GRCh37):g.51121756-51187704del	66	1	<i>De novo</i>	SHANK3 ; <i>ACR</i>	41	4	5.4	1.9-15.0	0.0013
18	SV	chrX(GRCh37):g.153295929-153296514del	0.6	1	<i>De novo</i>	MECP2 [‡]	41	0	22.6	1.4-367.6	0.04
31	CNV complex	chr4(GRCh37):g.183693432-183756173dup <i>Insertion point:</i> chrX(GRCh37):g.53318362-53318363	62 -	3 -	<i>De novo</i>	<i>TENM3</i> IQSEC2	5 28	0 0	0.6 7.6	0.2-2.4 1.0-56.1	1.0 0.092
37	CNV complex	chr3(GRCh37):g.48532000-49156000dup [†] chr3(GRCh37):g.49298000-49848000dup [†] chr3(GRCh37):g.49849505-49870969del chr3(GRCh37):g.49872000-49958000dup [†]	624 550 21 86	3 3 1 3	<i>De novo</i>	n=22 n=20 n=2 n=4	22	0	23.6	1.4-388.7	0.033
40	CNV	chr3(GRCh37):g.136003159delinsGATGTTTCA chr3(GRCh37):g.136003363-136385607del chr3(GRCh37):g.136385640-136385685del chr3(GRCh37):g.136385737-136385739del	- 382 - -	- 1	<i>De novo</i>	STAG1 ; <i>PCCB</i>	9 [¶]	0	10.0	0.6-171.1	0.26
48	SV	chrX(GRCh37):g.53424894-53427008del	2.1	1	<i>De novo</i>	SMC1A [§]	33	0	17.3	1.1-282.9	0.075
49	SV	chr1(GRCh37):g.40247181-40256104dup	9	3	<i>De novo</i>	BMP8B [§]	2	0	2.6	0.1-54.6	1.0
50	CNV	chr16(GRCh37):g.29567295-30177916 [‡]	611	1	<i>De novo</i>	n=29	31	4	4.1	1.4-11.5	0.0093
12	SV	chr8(GRCh37):g.100887349-100889133del	1.7	1	<i>Maternal</i>	VPS13B ^{§,‡}	5	0	5.8	0.3-104.2	0.80
	CNV	chr8(GRCh37):g.100147792-100270123del	122	1	<i>Paternal</i>	VPS13B [‡]	0	0			

Genes highlighted in bold are either listed as known ID genes or candidate ID genes. Please note that all patients had 250K SNP microarrays. Re-evaluation of these data showed that for all but one CNV the number of probes within the region was insufficient, either because of the small genomic size of the CNV, or due to uneven genome-wide probe spacing leaving fewer probes than required for the hidden Markov model algorithms to be identified.

* Primary method used to identify the rearrangement (see also Supplementary Methods).

† Not assessed at base-pair level due to complexity of CNV event including an inversion, duplication and deletion.

‡ Not assessed at base-pair level as the CNV event, involving a known microdeletion syndrome region, is mediated by low-copy repeats.

§ Single exon.

|| Number of genes affected rather than individual gene names are provided due to the large number of genes.

¶ Observed 13 times in Decipher.

Corrected for multiple testing using Benjamini–Hochberg with a false discovery rate (FDR) of 0.1.

☆ VPS13B recessive ID gene.

Extended Data Table 2 | *De novo* SNVs of potential clinical relevance identified using WGS

Trio	Gene	Protein effect	Mutation type	PhyloP [‡]	Gene Classification [§]
1	<i>NGFR</i>	p.(Cys122Arg)	Missense	4.97	-
2	<i>GFPT2</i>	p.(Thr680Ser)	Missense	6.02	-
6	<i>WWP2</i>	p.(Gly10Gly) [†]	Synonymous	-0.12	-
7	<i>TBR1</i>	p.(Gln373Arg)	Missense	3.51	Known
9	<i>WDR45</i>	p.(Cys344Alafs*67)	Frameshift		Known
13	<i>SMC1A</i>	p.(Asn788Lysfs*10)	Frameshift		Known
15	<i>SPTAN1</i>	p.(Glu91Lys)	Missense	5.69	Known
17	<i>ASUN</i>	p.(Gln99*)	Nonsense		-
21	<i>ALG13</i>	p.(Asn107Ser)	Missense	1.34	Known
21	<i>RAI1</i>	p.(Gln88*)	Nonsense		Known
22	<i>MED13L</i>	p.(Asp860Gly)	Missense	4.75	Candidate
24	<i>BRD3</i>	p.(Phe334Ser)	Missense	4.48	-
25	<i>SATB2</i>	p.(Gln310delinsHisCysLysAlaThr)	Insertion		Known
26	<i>PPP2R5D</i>	p.(Trp207Arg)	Missense	5.13	Candidate
27	<i>KCNA1</i>	p.(Thr371Ile)	Missense	5.69	Known
28	<i>SCN2A</i>	p.(Gln1521*)	Nonsense		Known
30	<i>MAST1</i>	p.(Pro1177Arg)	Missense	5.28	-
34	<i>APPL2</i>	p.(Ser329*)	Nonsense		-
41	<i>NACC1</i>	p.(Arg468Cys)	Missense	3.51	-
43	<i>POGZ</i>	p.(Arg1001*)	Nonsense		Candidate
46	<i>TBR1</i>	p.(Thr532Argfs*144)	Frameshift		Known
49	<i>KANSL2</i>	p.(Gly151Gly) [†]	Synonymous	1.58	Candidate

A dash indicates genes that have not yet been implicated in ID, but fulfil the criteria for diagnostic reporting of a pathogenic variant (that is, a possible cause for ID).

[†] Predicted effect on splicing.

[‡] PhyloP score for nonsense and frameshift mutations is not provided as this is deleterious mutations regardless of their evolutionary conservation.

[§] 'Known' refers to known ID gene whereas 'Candidate' refers to a gene that is listed on the candidate ID gene list.

^{||} Since the inclusion of this patient in this study, the same *de novo* mutation in *ALG13* has been described elsewhere¹⁶. This may suggest that this mutation, despite its low conservation and the identification of a nonsense mutation in *RAI1*, may also contribute to the disease phenotype in this patient. See also Supplementary Table 8 legend.

Engineering a memory with LTD and LTP

Sadegh Nabavi^{1*}, Rocky Fox^{1*}, Christophe D. Proulx¹, John Y. Lin², Roger Y. Tsien^{2,3} & Roberto Malinow¹

It has been proposed that memories are encoded by modification of synaptic strengths through cellular mechanisms such as long-term potentiation (LTP) and long-term depression (LTD)¹. However, the causal link between these synaptic processes and memory has been difficult to demonstrate². Here we show that fear conditioning^{3–8}, a type of associative memory, can be inactivated and reactivated by LTD and LTP, respectively. We began by conditioning an animal to associate a foot shock with optogenetic stimulation of auditory inputs targeting the amygdala, a brain region known to be essential for fear conditioning^{3–8}. Subsequent optogenetic delivery of LTD conditioning to the auditory input inactivates memory of the shock. Then subsequent optogenetic delivery of LTP conditioning to the auditory input reactivates memory of the shock. Thus, we have engineered inactivation and reactivation of a memory using LTD and LTP, supporting a causal link between these synaptic processes and memory.

To examine the relation between synaptic plasticity and memory, we used cued-fear conditioning^{3–8} in rats, wherein a neutral conditioned stimulus (CS), such as a tone, when paired with an aversive unconditioned stimulus (US), results in a tone-driven conditioned response (CR) indicating memory of the aversive stimulus. Temporally (but not non-temporally) pairing a tone with a shock led to a robust CR (reduced lever pressing to a previously learned cued lever-press task⁹; Extended Data Fig. 1) during subsequent testing with a tone alone^{3–8} (Fig. 1a). To investigate the synaptic basis underlying this associative memory, we replaced a tone with optogenetic stimulation of neural inputs to the lateral amygdala originating from auditory nuclei. We injected an adeno-associated virus (AAV) expressing a variant of the light-activated channel ChR2, oChIEF, that can respond faithfully to 50–100 Hz stimuli¹⁰, into the medial geniculate nucleus and auditory cortex (Extended Data Fig. 2). After the channel reached axonal terminals in the lateral amygdala (Extended Data Fig. 3), a cannula permitting light delivery was placed targeting the dorsal tip of the lateral amygdala (Extended Data Fig. 4). An optical CS alone (a 2 min 10 Hz train of 2 ms pulses, see Methods) had no effect on lever pressing (Extended Data Fig. 5). However, temporally (but not non-temporally) pairing the optical CS with a foot shock (see Methods) led to a CR (Fig. 1b) that was sensitive to extinction (see below) and blocked by NMDA receptor inhibition during conditioning (Extended Data Fig. 6), indicating the generation of an associative memory.

To examine if LTP occurred after pairing optical CS with foot shock^{3–8} (Fig. 1d), we prepared amygdala brain slices from animals receiving unpaired, paired or no conditioning, and measured the AMPA receptor component (A) and NMDA receptor component (N) of the optically driven synaptic response (Fig. 1c). The A/N ratio increased in animals receiving paired conditioning indicating that LTP had occurred^{11,12} at optically driven inputs to lateral amygdala neurons.

Can memories be inactivated? If LTP occurred at the optically driven synapse onto the lateral amygdala, and this LTP contributes to the memory, reversing LTP with LTD^{13,14} should inactivate the memory. Animals that displayed CR after paired optical CS-shock conditioning were exposed to an optical LTD protocol (see Methods). One day later, animals were tested with optical CS and displayed no CR, indicating inactivation of

the memory of the shock by LTD (Fig. 2a, b, f). Next we examined if memories can be reactivated. To these animals we delivered an optical LTP protocol (see Methods). One day later, animals displayed a CR

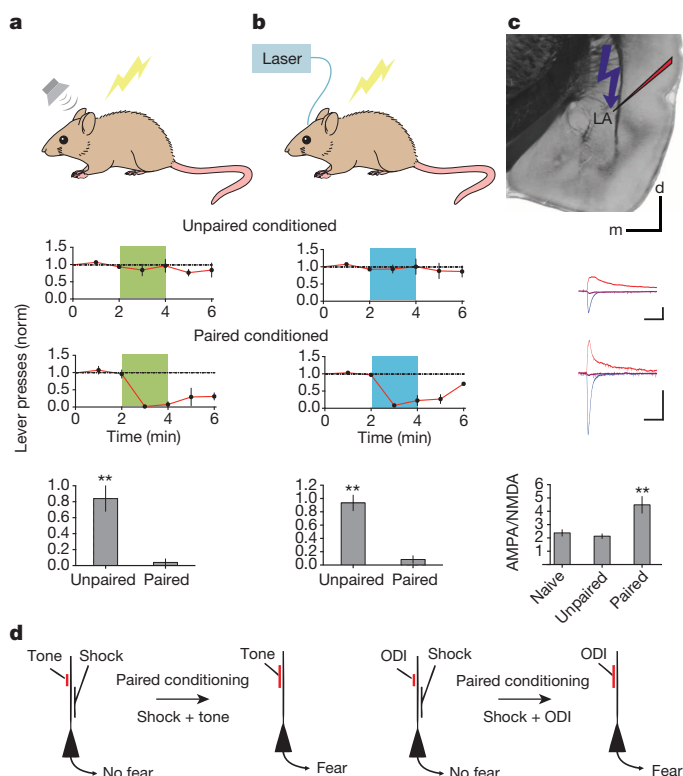


Figure 1 | Fear conditioning with tone or optogenetics. **a**, Top, diagram of rat receiving tone and shock during conditioning. Rats exposed to unpaired ($n = 5$, middle) or temporally paired ($n = 5$, bottom) tone and shock were tested one day later by a tone (green). Time plots show normalized number of lever presses (1 min bins) to a previously learned cued lever-press task. Bar graph shows normalized number of lever presses during the first minute of tone. **b**, Top, diagram of rat receiving optogenetically driven input (ODI) stimulation and shock during conditioning. Rats ($n = 8$) received unpaired (middle) and one day later temporally paired (bottom) ODI and shock. Time graphs as in **a**, except animals were tested by 10 Hz ODI (blue). Bar graph as in **a** for 10 Hz ODI. **c**, Top, experimental design; averaged optically driven synaptic responses obtained at -60 mV (blue), $+40$ mV (red) and 0 mV holding potential for cells from animals that received unpaired (top) or paired (bottom) conditioning. Traces were scaled to match NMDA-mediated currents. Bar graph plots average AMPA/NMDA (no conditioning, 2.4 ± 0.2 , $n = 11$ from 6 rats; unpaired conditioning 2.1 ± 0.2 , $n = 10$ from 6 rats; paired conditioning 4.4 ± 0.6 , $n = 8$ from 4 rats). Scale bars, 100 pA, 50 ms, 1 mm. **d**, Synaptic modification model. Temporally pairing of tone (left) or ODI (right) and shock inputs to lateral amygdala neurons leads to potentiation of tone (left) or ODI (right) input, which can contribute in triggering CR. Here and throughout: NS, not significant; * $P < 0.05$; ** $P < 0.01$; error bars, s.e.m. See Methods for details.

¹Center for Neural Circuits and Behavior, Department of Neuroscience and Section of Neurobiology, University of California at San Diego, California 92093, USA. ²Department of Pharmacology, University of California at San Diego, California 92093, USA. ³Howard Hughes Medical Institute, University of California at San Diego, California 92093, USA.

*These authors contributed equally to this work.

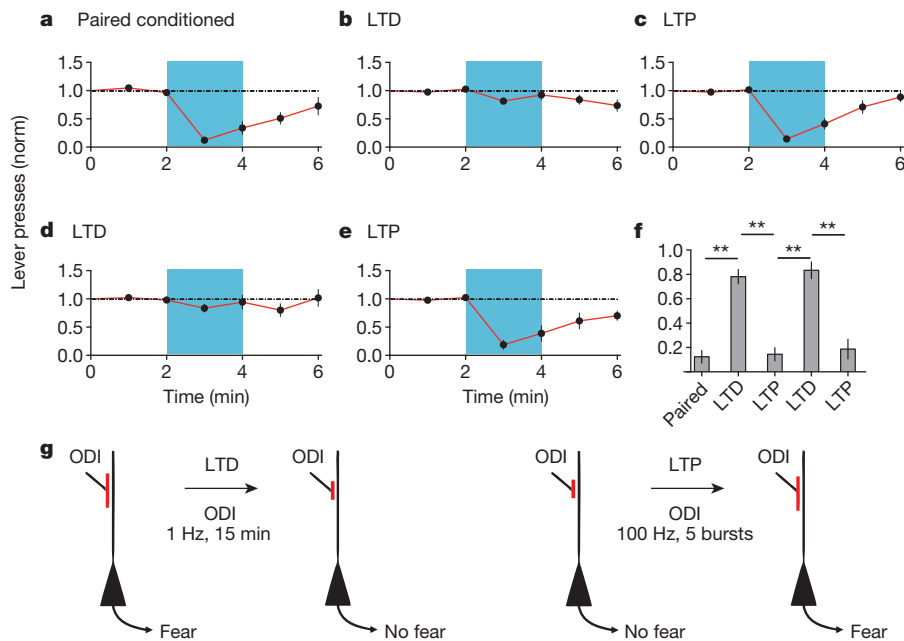


Figure 2 | LTD inactivates and LTP reactivates memory. **a–e**, A single group of rats ($n = 12$) was tested for CR two days following paired conditioning of ODI and shock (**a**). Graphs as in Fig. 1. After testing, animals were delivered an optical LTD protocol and tested for CR one day later (**b**). After testing, animals were delivered an optical LTP protocol and tested for CR one day later (**c**). After testing, animals were delivered another optical LTD protocol and tested for CR

one day later (**d**). After testing, animals were delivered another optical LTP and tested for CR one day later (**e**). **f**, Normalized lever presses one minute into optical CS after different protocols (as indicated). **g**, Cellular models of synaptic modifications occurring in the lateral amygdala that may contribute to behavioural responses following LTD (left) or LTP (right) protocols delivered to ODI.

(Fig. 2c, f), suggesting reactivation of the memory. Synapses are capable of undergoing multiple rounds of bidirectional plasticity¹³. We thus delivered a second optical LTD protocol; the next day animals produced no CR (Fig. 2d, f), indicating re-inactivation of the memory. Subsequent optical LTP conditioning recovered the CR (Fig. 2e, f and Extended Data Fig. 7) indicating reactivation of the memory. The behavioural effects of LTD and LTP conditioning were rapid and long lasting (Extended Data Fig. 8). These experiments suggest that a necessary component of the optical CS-triggered memory of the shock can be inactivated by LTD and reactivated by LTP.

In the experiments described above, LTP may be restoring a memory of the shock or merely potentiating random inputs that are sufficient to drive lateral amygdala neurons that produce fear and reduce lever pressing. Thus, we examined if generation of a CR by an LTP protocol requires prior optical CS-shock pairing. Indeed, an LTP protocol produced a CR only in animals that had previously received optical CS-shock conditioning (Fig. 3). These results support the view that LTP reactivates the memory that was formed by optical CS-shock pairing and inactivated by LTD.

To confirm that the test and conditioning stimuli were producing the expected synaptic effects, we conducted *in vivo* recordings in the lateral amygdala of anaesthetized rats expressing oChIEF in auditory regions (see Methods). Brief light pulses at the recording site produced *in vivo* field responses (that resembled optically evoked responses in amygdala brain slices; Extended Data Fig. 9), which were not affected by optical CS, depressed by optical LTD conditioning and potentiated by optical LTP conditioning (Fig. 4 and Extended Data Fig. 10). These results confirm that the synaptic stimulation conditioning protocols used to perturb behaviour modify synapses in the expected manner.

To examine further the relationship between these synaptic stimulation conditioning protocols and memory processes, we tested the effects of these protocols on auditory cued-fear conditioning. We first asked if optical LTD could inactivate tone-induced fear conditioning. In two groups of naive animals, we infected unilaterally auditory regions with AAV-oChIEF, and pharmacologically ablated the contralateral amygdala (see Methods). One group of animals received tone paired with shock,

which led to a tone-evoked CR (Fig. 5a, d). A second group of animals received the same tone paired with shock conditioning, immediately followed by an optical LTD protocol. This second group showed significantly reduced tone-evoked CR (Fig. 5b, d); subsequent tone conditioning without an optical LTD protocol produced a tone-evoked CR (Fig. 5c, d). This result is consistent with a memory model in which tone conditioning induces LTP at auditory inputs to the lateral amygdala and that subsequent LTD at these synapses reverses LTP and thereby inactivates the memory.

Next we examined extinction, a process whereby repeated exposure to a CS (in the absence of a shock (US)) leads to a reduced CR. We asked if optical LTP reverses extinction of tone conditioning. Animals received tone conditioning and an extinction protocol (see Methods), which removed the CR (Fig. 5e). Delivery of an optical LTP protocol did not restore the CR (Fig. 5f, g), consistent with the view that extinction is not a weakening of synapses potentiated during paired conditioning¹⁴. Similarly, animals receiving paired optical CS-shock conditioning produced a CR that could be removed by repeated exposure to optical CS (see Methods) and optical LTP did not recover the CR (Fig. 5h–k), again demonstrating that extinction is not LTD.

Prior studies examining the relation between LTP, LTD and memory have employed pharmacological (for example, ref. 15) or genetic (for example, refs 16, 17) manipulations to perturb and demonstrate parallels between cellular and behavioural processes. Other studies have measured randomly sampled sites in regions required for memory formation to detect changes in biochemistry and synaptic transmission following memory formation^{18–21}. However, selective perturbation of synapses that are employed to form a memory was not possible in these studies. Here by optogenetically isolating a neural input that can be used to form an associative memory, we can selectively manipulate synapses driven by this input and assess directly the relationship between cellular and behavioural processes.

Formation of an associative memory produced LTP at the lateral amygdala optogenetic input, as indicated by an increased A/N. Such LTP appears to be required as delivery of an LTD conditioning stimulus that can reverse LTP effectively removed the ability of the optogenetic

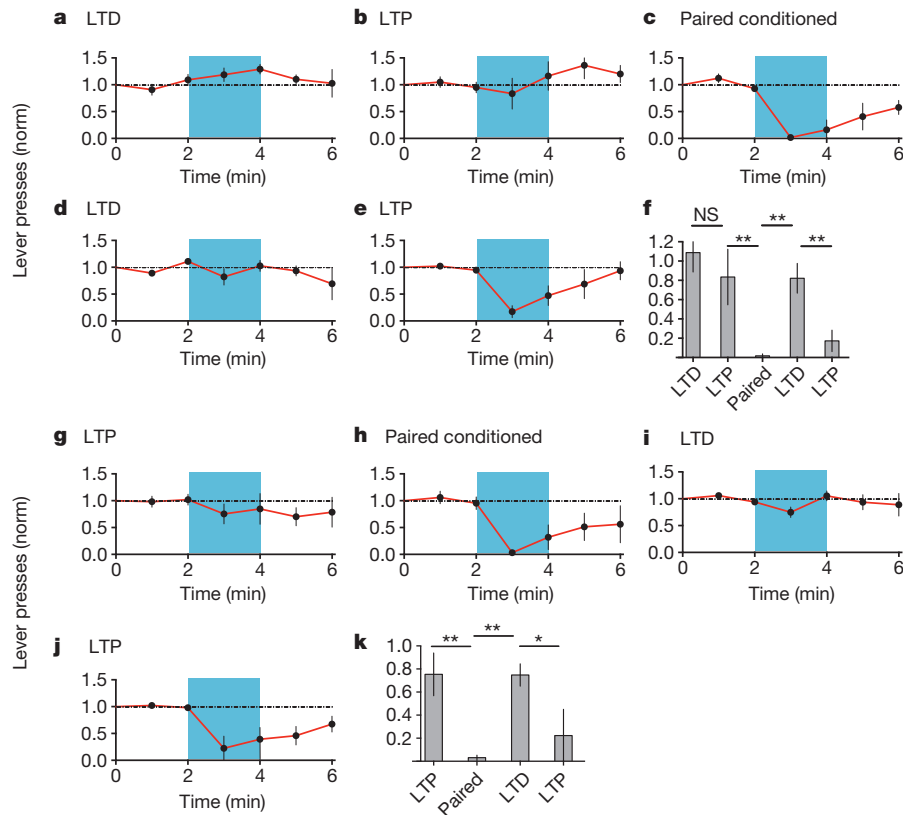


Figure 3 | LTP produces conditioned response only after prior paired conditioning. **a–f**, A naive group of animals ($n = 4$) was tested for CR one day after LTD protocol (**a**), one day after subsequent LTP protocol (**b**), one day after subsequent paired optical CS-shock conditioning (**c**), one day after subsequent LTD protocol (**d**) and one day after subsequent LTP protocol (**e**). **f**, Graph of normalized lever presses one minute into optical CS one day

input to elicit the memory. Furthermore, subsequent delivery of an LTP conditioning stimulus to the optogenetic neural input restored the CR. Our data support the view that LTP had reactivated the memory of the aversive stimulus, because delivery of an LTP protocol without prior formation of the memory did not evoke a CR. Our findings demonstrate that memories of aversive events formed through activation of selected

following indicated protocols. **g–k**, A separate naive group of animals ($n = 5$) was tested for CR one day after LTP protocol (**g**), one day after paired optical CS-shock conditioning (**h**), one day after subsequent LTD protocol (**i**) and one day after subsequent LTP protocol (**j**). **k**, Graph shows normalized lever presses one minute into optical CS one day following indicated protocols. Note that CR is seen following LTP protocol only after prior paired conditioning.

inputs can be turned off and on by conditioning protocols that produce bidirectional synaptic plasticity at those inputs, strengthening the causal relation between synaptic plasticity and memory formation²².

It is notable that optical LTP in naive animals did not produce a CR; whereas in these animals, optical LTP did produce a CR after optical CS-shock pairing and optical LTD. This result suggests that non-specific

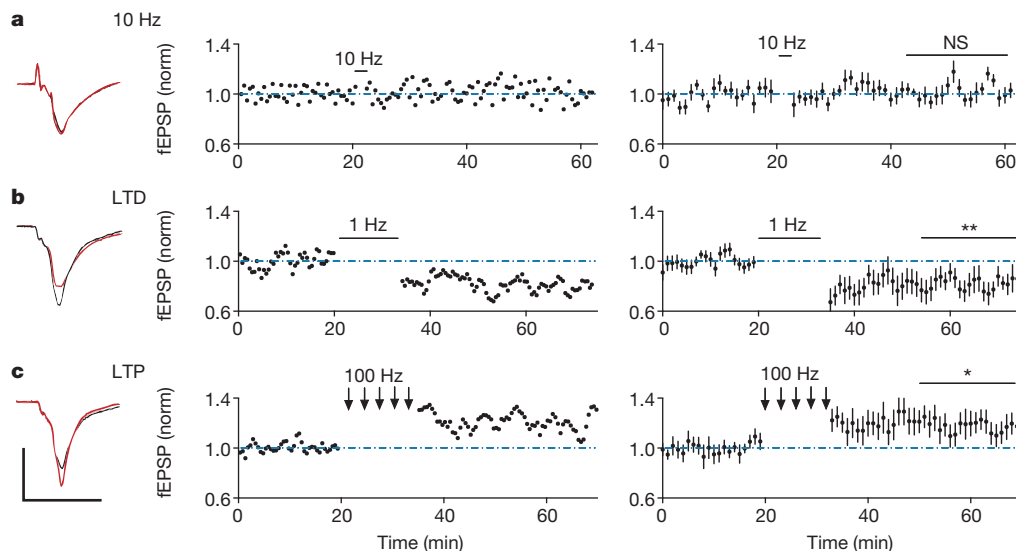


Figure 4 | In vivo electrophysiological responses to 10 Hz, LTD and LTP protocols. **a–c**, Left, *in vivo* field response (average of 20 responses) in lateral amygdala to single optical stimulus before (black) and after (red) indicated conditioning protocol. Plot of individual experiment (middle) or average of 10

experiments recorded from 10 rats (right) of field EPSP slope (normalized to baseline period) before and after indicated stimulation. Average baseline normalized value 30–40 min following conditioning: 10 Hz, $102.2 \pm 5\%$; 1 Hz, $82 \pm 8\%$; 100 Hz, $118 \pm 9\%$. Scale bars, 1 mV, 10 ms.

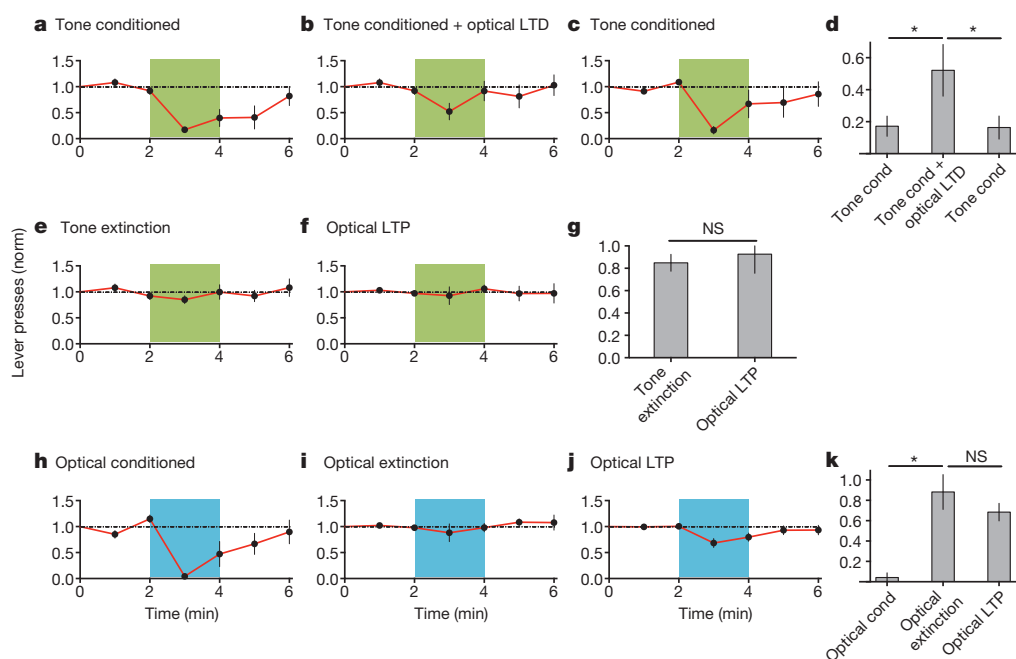


Figure 5 | Optical LTD protocol significantly reduces auditory fear conditioning; optical LTP does not reverse extinction. **a, b,** Separate groups of animals were exposed to paired tone and shock conditioning (**a**, $n = 5$) or paired tone and shock conditioning followed by optical LTD protocol (**b**, $n = 5$), and subsequently tested for CR with tone (green). **c,** Animals shown in **b** were subsequently exposed to paired tone and shock conditioning and tested for CR. **d,** Optical LTD significantly reduces auditory fear conditioning.

e–g, Animals shown in **c** were exposed to auditory extinction protocol and tested for CR (**e**), and subsequently received optical LTP and were tested for CR (**f**). **g,** Optical LTP did not reverse auditory extinction. **h–k,** A naive group of animals ($n = 5$) received paired optical conditioning and tested for CR (**h**); then received optical extinction protocol (see Methods) and were tested for CR (**i**); then received optical LTP protocol and were tested for CR (**j**). **k,** Optical LTP did not reverse optical extinction.

potentiation of auditory inputs to the lateral amygdala is not sufficient to produce a CR. It may be that specific potentiation onto a subset of inputs, presumably those neurons also activated by the foot shock, is necessary to produce a CR. Furthermore, the pairing of optical CS with shock probably produces additional modifications (not produced by optical LTP alone) that may be required to produce a CR^{23–26}. Thus, LTP at auditory inputs to the amygdala may be necessary but not sufficient to produce an associative memory.

Our studies complement recent studies that have used optogenetics to examine how neuronal assemblies can represent a memory^{27–29}. In those studies synaptic mechanisms were not examined. Our studies suggest that LTP is used to form neuronal assemblies that represent a memory. Furthermore, our findings predict that LTD could be used to disassemble them and thereby inactivate a memory.

METHODS SUMMARY

Surgery. AAV expressing a variant of the light-activated channel ChR2, oChIEF¹⁰, was injected into the auditory nuclei of 6–8-week-old rats. Then 3–4 weeks later an optic fibre cannula was placed above the dorsal tip of the lateral amygdala (dLA). **Behaviour.** Rats were trained to associate lever press for a reward and tested for a CR during the lever press task. Tone conditioning protocol consisted of 10 pairs of 20 s tones co-terminated with 500 ms of 0.5 mA foot shock. Optical conditioning was as above, except that each tone was replaced with 1 s of 10 Hz blue light.

Optical plasticity induction. LTD was induced with 900 2 ms pulses of light delivered at 1 Hz. LTP was induced with 5 trains of light, each train containing 100 2 ms pulses, delivered at 100 Hz, with 3 min inter-train intervals.

During all behavioural manipulations the light intensity remained the same for each animal.

In vitro recording. Acute slices were prepared from rats expressing AAV-oChIEF in the auditory nuclei. Extracellular field potentials (fEPSPs) or excitatory postsynaptic current (EPSC) responses were obtained from the dLA by optical stimulation of the auditory projections.

In vivo recording. Rats expressing AAV-oChIEF in auditory nuclei were anaesthetized and a recording glass pipette was placed in the dLA. fEPSPs were evoked using an optic fibre placed above the recording site.

Analysis. A CR was measured as the reduction in the frequency of lever presses during the CS (2 min of tone or 10 Hz light stimulation).

fEPSP initial slope and EPSC amplitude were measured.

All values indicate mean \pm s.e.m. Student's paired and non-paired *t*-tests were used with $P < 0.05$ considered as significant. All behavioural data were reanalysed with Wilcoxon rank-sum test which produced similar significance values as the *t*-test.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 August 2013; accepted 27 March 2014.

Published online 1 June 2014.

1. Squire, L. R. & Kandel, E. R. *Memory: From Mind to Molecules* 2nd edn (Roberts & Co., 2009).
2. Stevens, C. F. A million dollar question: does LTP = memory? *Neuron* **20**, 1–2 (1998).
3. LeDoux, J. E. Emotion circuits in the brain. *Annu. Rev. Neurosci.* **23**, 155–184 (2000).
4. Fanselow, M. S. & Poulos, A. M. The neuroscience of mammalian associative learning. *Annu. Rev. Psychol.* **56**, 207–234 (2005).
5. Maren, S. & Quirk, G. J. Neuronal signalling of fear memory. *Nature Rev. Neurosci.* **5**, 844–852 (2004).
6. Davis, M. The role of the amygdala in fear and anxiety. *Annu. Rev. Neurosci.* **15**, 353–375 (1992).
7. Sah, P., Westbrook, R. F. & Luthi, A. Fear conditioning and long-term potentiation in the amygdala: what really is the connection? *Ann. NY Acad. Sci.* **1129**, 88–95 (2008).
8. Pape, H. C. & Pare, D. Plastic synaptic networks of the amygdala for the acquisition, expression, and extinction of conditioned fear. *Physiol. Rev.* **90**, 419–463 (2010).
9. Repa, J. C. *et al.* Two different lateral amygdala cell populations contribute to the initiation and storage of memory. *Nature Neurosci.* **4**, 724–731 (2001).
10. Lin, J. Y., Lin, M. Z., Steinbach, P. & Tsien, R. Y. Characterization of engineered channelrhodopsin variants with improved properties and kinetics. *Biophys. J.* **96**, 1803–1814 (2009).
11. Muller, D., Joly, M. & Lynch, G. Contributions of quisqualate and NMDA receptors to the induction and expression of LTP. *Science* **242**, 1694–1697 (1988).
12. Kauer, J. A., Malenka, R. C. & Nicoll, R. A. A persistent postsynaptic modification mediates long-term potentiation in the hippocampus. *Neuron* **1**, 911–917 (1988).

13. Mulkey, R. M. & Malenka, R. C. Mechanisms underlying induction of homosynaptic long-term depression in area CA1 of the hippocampus. *Neuron* **9**, 967–975 (1992).
14. Herry, C. *et al.* Neuronal circuits of fear extinction. *Eur. J. Neurosci.* **31**, 599–612 (2010).
15. Morris, R. G., Anderson, E., Lynch, G. S. & Baudry, M. Selective impairment of learning and blockade of long-term potentiation by an *N*-methyl-D-aspartate receptor antagonist, AP5. *Nature* **319**, 774–776 (1986).
16. Silva, A. J., Paylor, R., Wehner, J. M. & Tonegawa, S. Impaired spatial learning in alpha-calcium-calmodulin kinase II mutant mice. *Science* **257**, 206–211 (1992).
17. Silva, A. J., Stevens, C. F., Tonegawa, S. & Wang, Y. Deficient hippocampal long-term potentiation in alpha-calcium-calmodulin kinase II mutant mice. *Science* **257**, 201–206 (1992).
18. Whitlock, J. R., Heynen, A. J., Shuler, M. G. & Bear, M. F. Learning induces long-term potentiation in the hippocampus. *Science* **313**, 1093–1097 (2006).
19. Rumpel, S., Ledoux, J., Zador, A. & Malinow, R. Postsynaptic receptor trafficking underlying a form of associative learning. *Science* **308**, 83–88 (2005).
20. Rogan, M. T., Stäubli, U. V. & LeDoux, J. E. Fear conditioning induces associative long-term potentiation in the amygdala. *Nature* **390**, 604–607 (1997).
21. McKernan, M. G. & Shinnick-Gallagher, P. Fear conditioning induces a lasting potentiation of synaptic currents *in vitro*. *Nature* **390**, 607–611 (1997).
22. Martin, S. J. & Morris, R. G. M. New life in an old idea: the synaptic plasticity and memory hypothesis revisited. *Hippocampus* **12**, 609–636 (2002).
23. Li, H. *et al.* Experience-dependent modification of a central amygdala fear circuit. *Nature Neurosci.* **16**, 332–339 (2013).
24. Paré, D., Quirk, G. J. & LeDoux, J. E. New vistas on amygdala networks in conditioned fear. *J. Neurophysiol.* **92**, 1–9 (2004).
25. Wilensky, A. E., Schafe, G. E., Kristensen, M. P. & LeDoux, J. E. Rethinking the fear circuit: the central nucleus of the amygdala is required for the acquisition, consolidation, and expression of Pavlovian fear conditioning. *J. Neurosci.* **26**, 12387–12396 (2006).
26. Ciocchi, S. *et al.* Encoding of conditioned fear in central amygdala inhibitory circuits. *Nature* **468**, 277–282 (2010).
27. Liu, X. *et al.* Optogenetic stimulation of a hippocampal engram activates fear memory recall. *Nature* **484**, 381–385 (2012).
28. Garner, A. R. *et al.* Generation of a synthetic memory trace. *Science* **335**, 1513–1516 (2012).
29. Ramirez, S. *et al.* Creating a false memory in the hippocampus. *Science* **341**, 387–391 (2013).

Acknowledgements We thank J. Isaacson, L. Squire and members of the Malinow laboratory for suggestions. This study was supported by NIH MH049159 and Cure Alzheimer's Foundation grants to R.M. and NIH grant NS27177 to R.T.; R.T. is an Investigator of the HHMI.

Author Contributions S.N. and R.M. designed the experiments and wrote the manuscript. S.N., R.F. and R.M. analysed the data. S.N., R.F. and C.D.P. performed the experiments. J.Y.L. and R.Y.T. provided the oChIEF-tdTomato construct.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.M. (rmalinow@ucsd.edu).

METHODS

Subject. Male Sprague-Dawley rats, aged 6–8 weeks for virus injection and cannula placement and 10–12 weeks for behavioural and electrophysiological studies, were housed two per cage and kept on a 12/12 h light–dark cycle (lights on/off at 7:00/19:00). The behavioural studies were done during daylight. All procedures involving animals were approved by the Institutional Animal Care and Use Committees of the University of California, San Diego.

Virus. We used a ChR variant, named oChIEF, which is a mammalian codon optimized version of ChIEF^{10,30} with the same properties except that it has stronger expression in mammalian cells and has an additional N-terminal amino acid residue. Expression was driven by the neuron-specific synapsin promoter³⁰.

Surgery. Male Sprague-Dawley rats, aged 6–8 weeks, were anaesthetized with isoflurane for stereotaxic injection of AAV-oChIEF into the medial geniculate nucleus (AP: –5.1 mm and –5.7 mm; ML: 2.9 mm; DV: –5.5 to –6.7 mm) and the auditory cortex (AP: –5.7 mm; ML: 4.8 mm with a 20° angle; DV: –4.5 to –5.7 mm). A total of 0.4–0.5 µl of virus was injected over a 10–15 min period. At the end of the injection, the pipet remained at the site for 5 min to allow for diffusion of the virus. An optic fibre cannula (Doric Lenses) was implanted just above the dorsal tip of the lateral amygdala (AP: –3.3 to –3.5 mm; ML: 4.2 mm; DV: –7 mm with a 7° angle) and secured to the skull with screws and dental cement. Rats were injected with 5 mg per kg carprofen (NSAID) after surgery.

Excitotoxic lesion. Rats aged 6–8 weeks were anaesthetized with isoflurane for stereotaxic injection of *N*-methyl-D-aspartate (NMDA) into one amygdala (AP: –3 mm; ML: 4.2 mm; DV: –7 to –8 mm with a 7° angle). 0.5 µl of NMDA (20 mg ml^{–1}) was injected over a 10–15 min period³¹. At the end of the injection, the pipet remained at the site for 5 min to allow for diffusion of the solution.

Behavioural assays

Training. Rats were trained to associate lever press for a reward (40 µl of 10% sucrose per lever press). During the training period rats were kept on a restricted water schedule (2 h daily of water *ad libitum*). Training context was a modular operant test chamber (12.5 × 10 × 13 inches) with a stainless grid floor and open roof located in a sound attenuating cubicle (Med Associates, St. Albans, VT). The test chamber was equipped with a retractable response lever, a liquid dispenser receptacle and a light above the dispenser that signalled when liquid was injected into the dispenser. The consumption of liquid was detected by a head entry detector in the receptacle; each successive liquid reward was subsequently followed with a 15 s delay after head removal from the receptacle. The system was controlled and the data collected through a MED-SYST-16 interface, which was controlled by MED-PCR IV software running on a PC. Rats were initially trained to associate the reward with the light above the dispenser receptacle. In a 45 min session, rats with at least 60 head entries into the receptacle were selected for lever press training.

Lever-press training was conducted in the same context as above, but this time rats had to press a lever to receive the liquid. The lever press turned the light above the receptacle on, which in the previous training session they had associated with liquid in the receptacle. Rats with a minimum of 6 responses per min in the first 10 min of the training session were selected for conditioning.

Tone conditioning. The conditioning chamber was a box (12 × 10.5 × 13 inches) with an electrified grid floor (Coulbourn Instruments, Allentown, PA) within a larger sound-attenuating box. Rats had full access to water 24 h before conditioning. Conditioning protocol consisted of 10 trials of 20 s tone (tone volume 80 dB), with randomized intervals (average interval duration 3 min). In the paired group tones were co-terminated with a 0.5 s 0.5 mA footshock (or a single 20 s tone co-terminated with a 1 s 0.5 mA footshock for mild conditioning, Fig. 5). In unpaired group tones and shocks were separated by at least 1 min. Paired and unpaired groups received equal number of tones (CS) and shocks (US) in the same context; however, only in the paired group did tone and shock coincide. The next day, conditioned rats were placed into the test chamber to measure the effect of CS on their lever presses (for details, see the section on testing).

Optical conditioning. Rats were placed into the conditioning chamber and were attached to an optic fibre patch cord connected to a 473 nm solid-state laser diode (OEM Laser Systems) with 15–20 mW of output from the 200 µm fibre. They were allowed to explore the chamber for 3 min before the conditioning. Optical conditioning was 10-trains of blue light (10 pulses of 10 Hz, 2 ms duration) applied at randomized intervals with an average of 3 min apart. For paired conditioning, the light stimulus co-terminated with 0.5 s of 0.5 mA footshock; in unpaired conditioning, the light and shock were separated by a minimum of 1 min. Paired and unpaired groups received equal number of light stimuli and shocks in the same context; however, only in the paired group did light and shock coincide. The delivery of shock and light was controlled by a pulse generator (Master-8; AMPI, Jerusalem, Israel). After the conditioning rats remained in the box for additional three minutes before returning to their home cage.

Testing. After the conditioning, rats were water restricted for 24 h before they were tested for lever press. Testing was done in the same context as training except that

the floor was a plastic sheet with white and red strips. Testing was a 7 min session in which rats had to press a lever to receive the liquid (10% sucrose). Rats were attached to the optic fibre patch cord, placed into the chamber, and allowed to explore the environment for 5 min before having access to the lever. The testing session, in which rats had free access to the lever, was a 3 min period of no light, followed by two minutes of light on (10 Hz of pulses with 2 ms duration), and 2 min of no light. At the end of the session rats were returned to their home cage. Only rats that in two consecutive days showed consistent reduction (>30%) in the lever press during the light-on period were used for further behavioural phases. Those which failed the test were examined histologically to locate the position of cannula and viral injection (Extended Data Fig. 4).

Tone-conditioned rats were tested in the same way except that they received 2 min of tone instead of light stimulation.

LTD induction. Within one hour following testing, rats were placed in a separate context, a translucent plastic container (22.5 × 15 × 12 inches), attached to the optic fibre patch cord and allowed to explore the environment for 3 min before LTD induction. Optical LTD was induced with 900 pulses of light, each 2 ms, at 1 Hz. After the induction rats remained in the chamber for 3 additional minutes before returning to their home cage.

LTP induction. Within one hour following testing, rats were placed in a separate context, a cardboard box (20.5 × 15.5 × 14.5 inches), attached to the optic fibre patch cord and allowed to explore the environment for 3 min before LTP induction. Optical LTP was induced with 5 trains of light (each train 100 pulses, 100 Hz) at 3 min inter-train intervals. After the induction, rats remained in the chamber for 3 additional minutes before returning to their home cage.

During all behavioural assays the light intensity remained the same for each animal. At the end of the experiment, animals were perfused and the location of the optic fibre was verified.

Systemic injection of MK801. Rats were anaesthetized with isoflurane for 5 min before being given an intraperitoneal injection of MK801 (ref. 32) (0.2 mg per kg) in sterile saline. The conditioning protocol was administered 30 min following injection.

Perfusion, slicing and imaging. Prior to perfusion, rats were administered a ketamine/dexdomitor (75 and 5 mg per kg respectively) mixture by intraperitoneal injection. Rats were then transcardially perfused with ~150 ml of saline followed by ~150 ml of 4% paraformaldehyde in 0.1 M phosphate buffer solution (PB, pH 7.4). Brains were then fixed overnight in the same solution and rinsed and stored in 0.1 M PB for slicing.

Brains were sliced coronally in 150 µm sections using a vibratome sectioning system and stored in PB. Slices were imaged using an Olympus MVX10 epifluorescent microscope to verify AAV-oChIEF-tdTomato expression in the MGN, auditory cortex, and their projections to the dorsal lateral amygdala. Additionally, appropriate positioning of the optic fibre cannula over the lateral amygdala was verified.

In vitro recording. For extracellular field potential recordings, acute slices (as described in ref. 33) were prepared from 3–4-month-old rats expressing AAV-oChIEF in the medial geniculate nucleus and/or auditory cortex. Extracellular field potentials were recorded with Axopatch-1D amplifiers (Axon Instruments) in dorsal tip of the lateral amygdala with glass electrodes (1–2 MΩ) filled with the perfusion solution. The auditory projection to the lateral amygdala was evoked by optical stimulation above the recording site. To measure AMPA-R field potential, 2,3-dihydroxy-6-nitro-7-sulfamoyl-benzo[f]quinoxaline-2,3-dione (NBQX) (10 µM) was added at the end of the experiments. Data were acquired and analysed using custom software written in Igor Pro (Wavemetrics). The perfusion solution contained: 119 mM NaCl, 2.5 mM KCl, 2 mM CaCl₂, 1 mM MgCl₂, 26 mM NaHCO₃, 1 mM NaH₂PO₄, 11 mM glucose (pH 7.4), and gassed with 5% CO₂/95% O₂.

For whole-cell recording, acute slices (as described in refs 34–36) were prepared from 3–4-month-old rats expressing AAV-oChIEF in the medial geniculate nucleus and/or auditory cortex. Whole-cell recordings were obtained from individual cells in dorsal tip of the lateral amygdala using glass pipettes (3–4 MΩ) filled with internal solution containing, in mM, cesium methanesulfonate 115, CsCl 20, HEPES 10, MgCl₂ 2.5, Na₂ATP 4, Na₃GTP 0.4, sodium phosphocreatine 10, and EGTA 0.6, at pH 7.25. External perfusion consisted of artificial cerebrospinal fluid (ACSF), containing 119 mM NaCl, 2.5 mM KCl, 26 mM NaHCO₃, 1 mM NaH₂PO₄, 11 mM glucose, supplemented with 1 mM MgCl₂, and 2 mM CaCl₂, 100 µM picrotoxin and 1 mM Sodium L-ascorbate. Synaptic responses were evoked every 10 s by stimulating auditory projections to the lateral amygdala using 2 ms of blue light generated by the epifluorescence microscope and passed through the ×60 objective lenses placed immediately above the recorded cell. The AMPA/NMDA ratio was calculated as the ratio of peak current at –60 mV to the current at +40 mV, 50 ms after stimulus; both values subtracted from the current at 0 mV.

In vivo recording. Four weeks after injection of AAV-oChIEF-tdTomato into auditory regions (8 animals were injected in both MGN and auditory cortex; 2 animals were injected in only the auditory cortex; results were pooled), rats were anaesthetized with a set of three injections of 700 µl urethane (330 mg ml^{–1}) given at 10 min

intervals 2 h before the recording³⁷ and then mounted on a custom-made stereotaxic frame with an adjustable angle, to hold the head in a fixed position during the recording. The body temperature was regulated by a heating pad. Using aseptic surgical tools the skull was exposed and a hole (~3 mm) was made, centred at -3.3 mm AP and 4.2 mm ML. The recording electrode was a glass pipet (4–5 M Ω) filled with 0.9% NaCl. The recording electrode was connected to a Axopatch-1D amplifier. The signal was amplified ($\times 1,000$), filtered (2K Hz) and digitized at 10 kHz using an Instrutech A/D interface. Data were acquired and analysed using custom software written in Igor Pro (Wavemetrics).

For optical stimulation, the optic fibre was glued to the glass pipet so that the tip of the fibre was 500 μ m above the tip of the glass pipet to form an optrode. The optic fibre was connected to a 473 nm solid-state laser diode (OEM Laser Systems). The parameters for the optical stimulation were identical to those used during behaviour (2 ms duration, 15–20 mW intensity). The optrode was slowly lowered in at a 7° angle following the start of stimulation. After establishing a stable baseline of at least 30 min (stimulation frequency 0.033 Hz) at the recording site (DV: -7 to -7.5), 2 min of 10 Hz stimulation was evoked, which was followed by 40 min of 0.033 Hz stimulation. Subsequent LTD and LTP, with the same parameters used in the behavioural assay, were induced 40 min apart. Electrode resistance and light intensity were monitored before and immediately after the recordings to ensure that there was no change in the course of recording. All animals were perfused after the recordings and the position of the recording site verified.

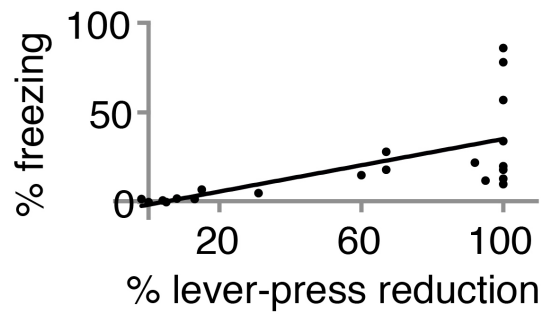
Analysis. The number of lever presses were binned for each minute and normalized to the 2-min period before light stimulation. Suppression ratio was measured by dividing the number of lever presses during the first minute of conditioning stimulus (tone or optical stimulation) by that immediately preceding the stimulus.

To minimize the voltage dependent conductance component, the initial slope of field excitatory postsynaptic potentials¹³ were measured using a custom written MATLAB program.

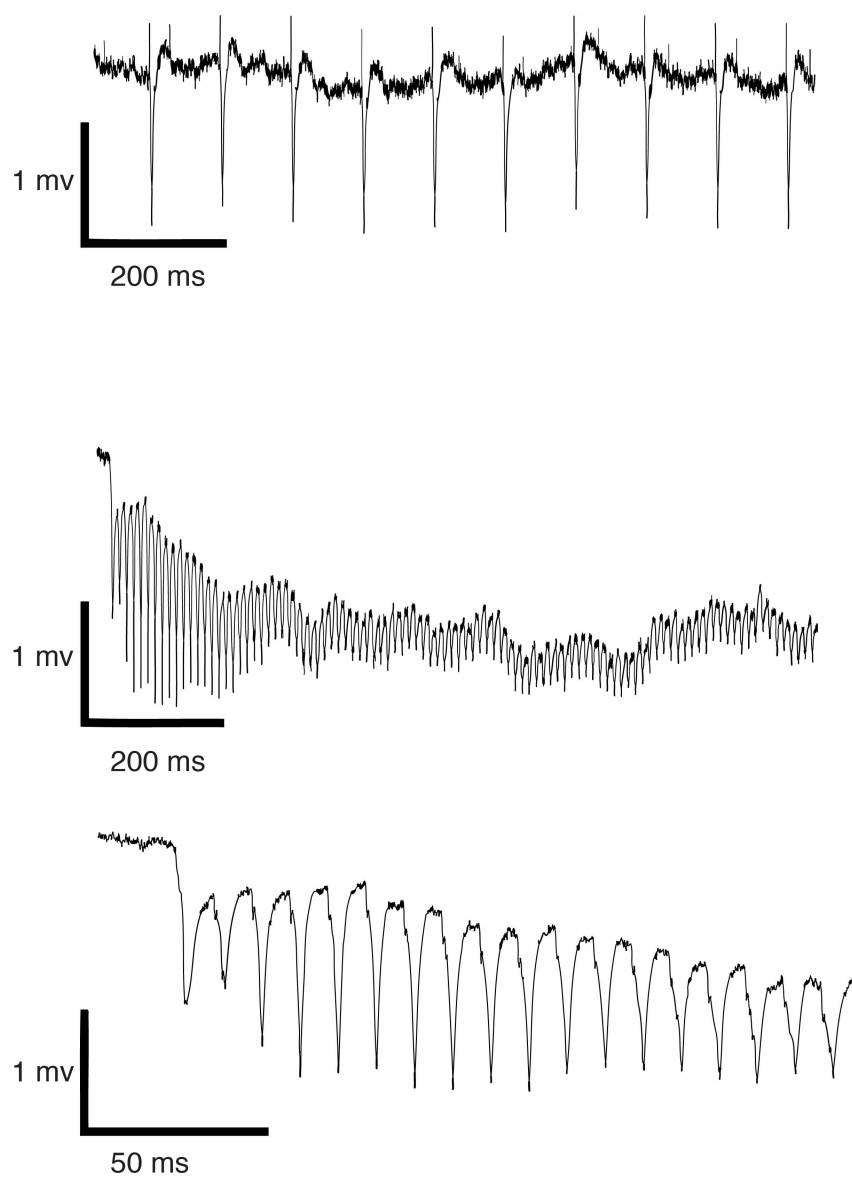
Excitatory postsynaptic current amplitude was measured by averaging a fixed 3 ms window covering the peak amplitude and subtracting from an average current window before stimulation.

All values given in the text and figures indicate mean \pm s.e.m. Student's paired and non-paired *t*-tests were used with $P < 0.05$ considered as significant. All behavioural data were also analysed with the Wilcoxon rank-sum test (MATLAB statistic toolbox) and yielded the same significance values as the *t*-test.

30. Lin, J. Y., Knutsen, P. M., Muller, A., Kleinfeld, D. & Tsien, R. Y. ReaChR: a red-shifted variant of channelrhodopsin enables deep transcranial optogenetic excitation. *Nature Neurosci.* **16**, 1499–1508 (2013).
31. Kirby, E. D., Jensen, K., Goosens, K. A. & Kaufer, D. Stereotaxic surgery for excitotoxic lesion of specific brain areas in the adult rat. *JoVE* **65**, e4079 (2012).
32. Li, X. F., Stutzmann, G. E. & LeDoux, J. E. Convergent but temporally separated inputs to lateral amygdala neurons from the auditory thalamus and auditory cortex use different postsynaptic receptors: *in vivo* intracellular and extracellular recordings in fear conditioning pathways. *Learn Mem.* **3**, 229–242 (1996).
33. Rumpel, S., LeDoux, J., Zador, A. & Malinow, R. Postsynaptic receptor trafficking underlying a form of associative learning. *Science* **308**, 83–88 (2005).
34. Li, B. *et al.* Synaptic potentiation onto habenula neurons in the learned helplessness model of depression. *Nature* **470**, 535–539 (2011).
35. Hoffman, D. A. & Johnston, D. Downregulation of transient K⁺ channels in dendrites of hippocampal CA1 pyramidal neurons by activation of PKA and PKC. *J. Neurosci.* **18**, 3521–3528 (1998).
36. Chen, B. T. *et al.* Cocaine but not natural reward self-administration nor passive cocaine infusion produces persistent LTP in the VTA. *Neuron* **59**, 288–297 (2008).
37. Rogan, M. T. & LeDoux, J. E. LTP is accompanied by commensurate enhancement of auditory-evoked responses in a fear conditioning circuit. *Neuron* **15**, 127–136 (1995).

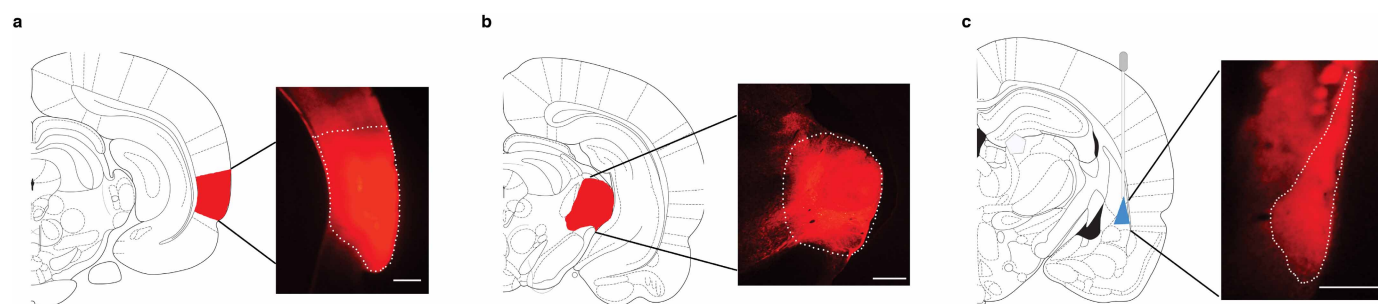


Extended Data Figure 1 | Freezing correlates well with reduction in lever presses to previously learned task. Plot of per cent freezing versus per cent reduction in lever presses to previously learned task. Best fit line indicates significant positive correlation ($R^2 = 0.4$; $P < 0.01$; F test). Data includes results from 3 manipulations (paired optical CS-shock conditioning, optical LTD and optical LTP). The per cent change in lever presses to previously learned task ($60\% \pm 9\%$) was significantly greater than change in per cent freezing ($20\% \pm 5\%$; $n = 21$; $P < 0.001$, paired Student's t -test).



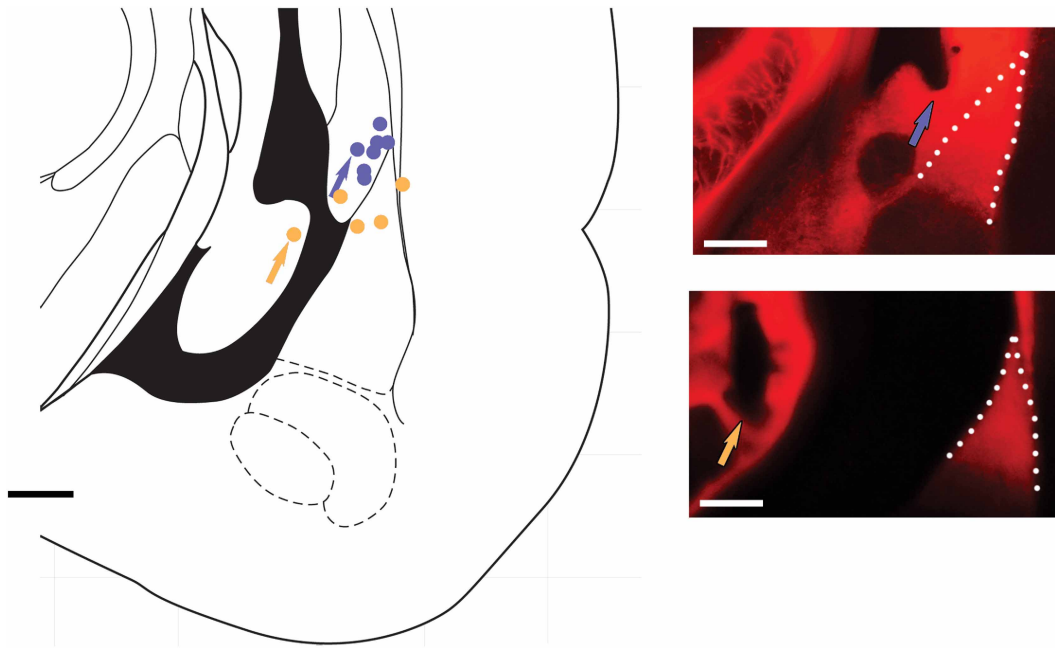
Extended Data Figure 2 | *In vivo* optically evoked synaptic responses in lateral amygdala. Field responses to 10 Hz (top) and 100 Hz optical stimulation (middle, bottom), obtained from animal infected with

AAV-oChIEF in auditory regions four weeks before recording. Note that the responses follow stimulation faithfully.



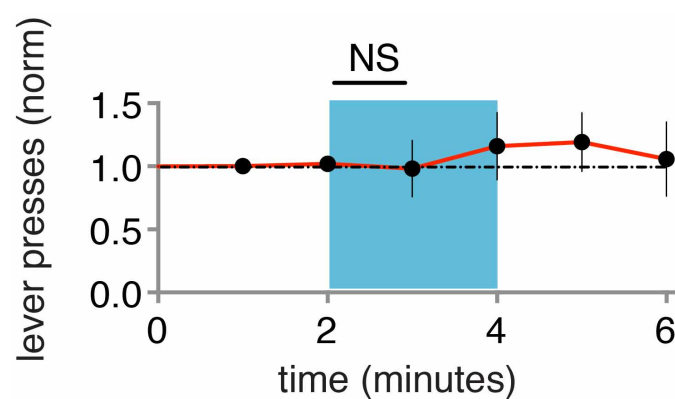
Extended Data Figure 3 | Expression of oChIEF in auditory regions reaches lateral amygdala. **a, b,** Diagram (left) and epifluorescent image (right) of coronal section of rat brain indicating areas expressing AAV-oChIEF-tdTomato 3–4 weeks after *in vivo* injection in auditory cortex (**a**) and medial

geniculate nucleus (**b**). **c,** Axonal expression of AAV-oChIEF-tdTomato in lateral amygdala (dashed white line); approximate placement of cannula and light (blue) indicated. Scale bars, 500 μ m.

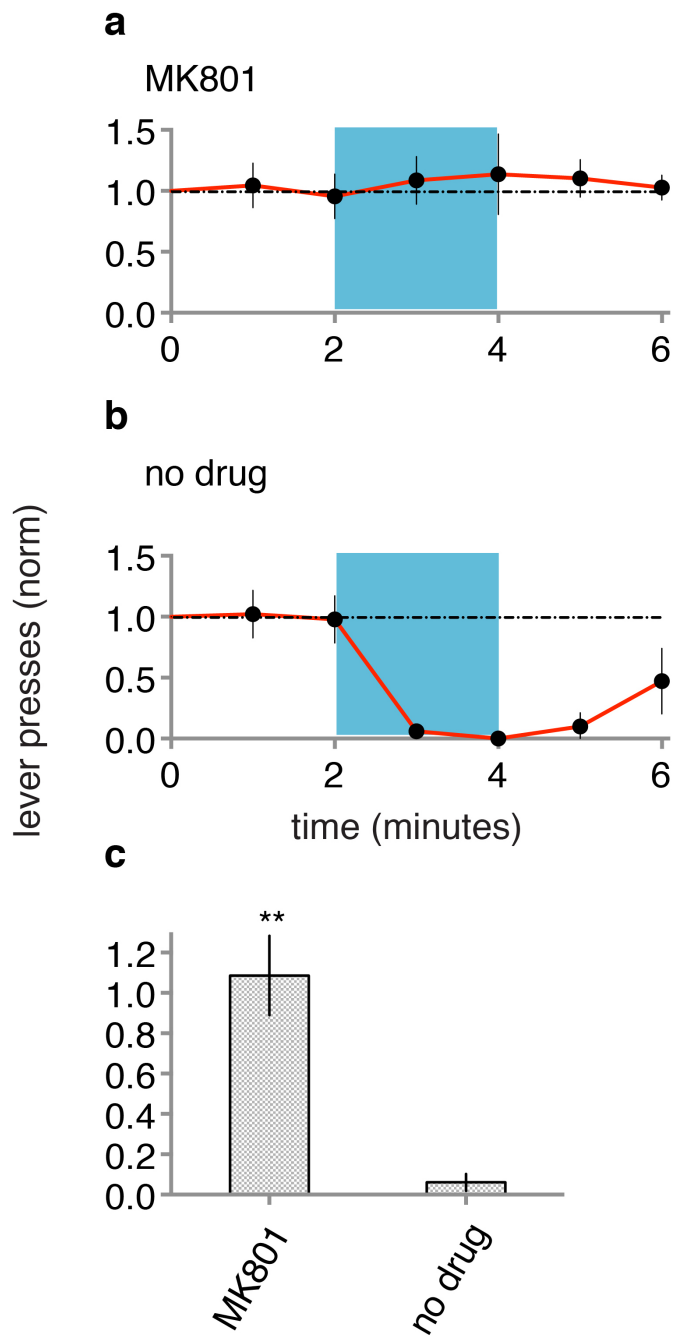


Extended Data Figure 4 | Optic fibre locations in representative group of rats used in the behavioural assays. Histologically assessed optic fibre tip location for rats which responded (blue circles; upper panel, right, is one example) or did not respond (orange circles; lower panel, right, is one example)

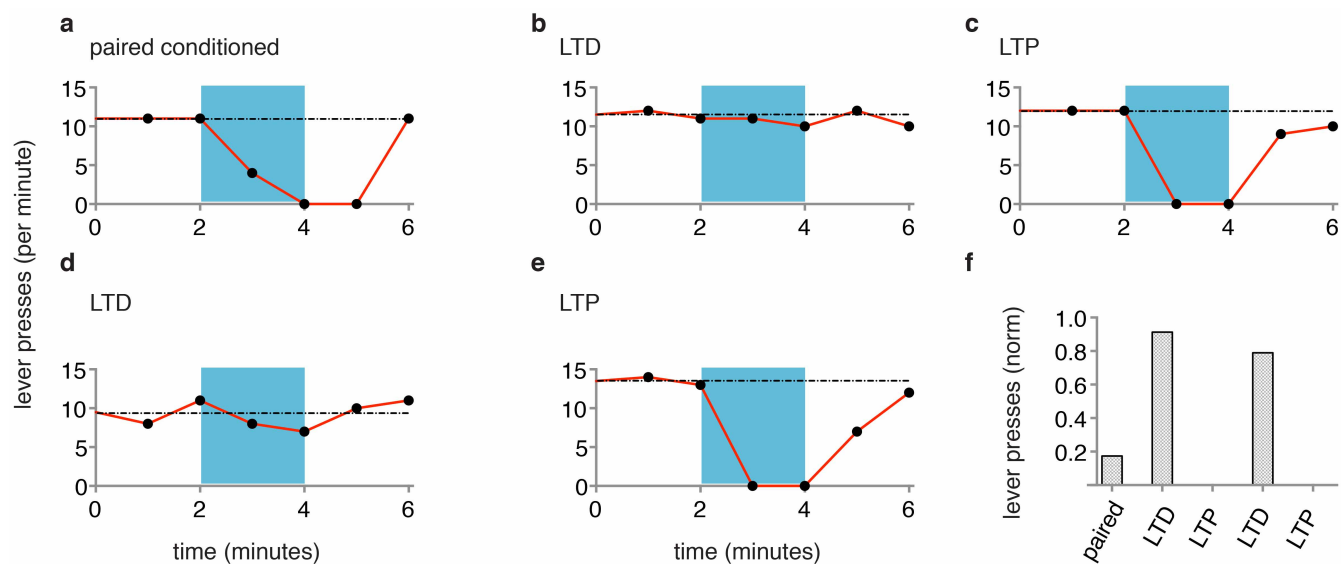
to optical conditioning. The arrow on the panels shows the location of the tip of optic fibre. Lateral amygdala is indicated by dashed line. Note that the ventricle opened during tissue sectioning in the lower image. Scale bars, 500 μm .



Extended Data Figure 5 | The 10 Hz test protocol does not produce CR. Test for CR (blue) in naïve animals ($n = 8$), as measured by changes in lever presses normalized to baseline period. Subsequent delivery of paired optical CS and shock produced CR in these animals (not shown). Each point represents data collected over 1 min.



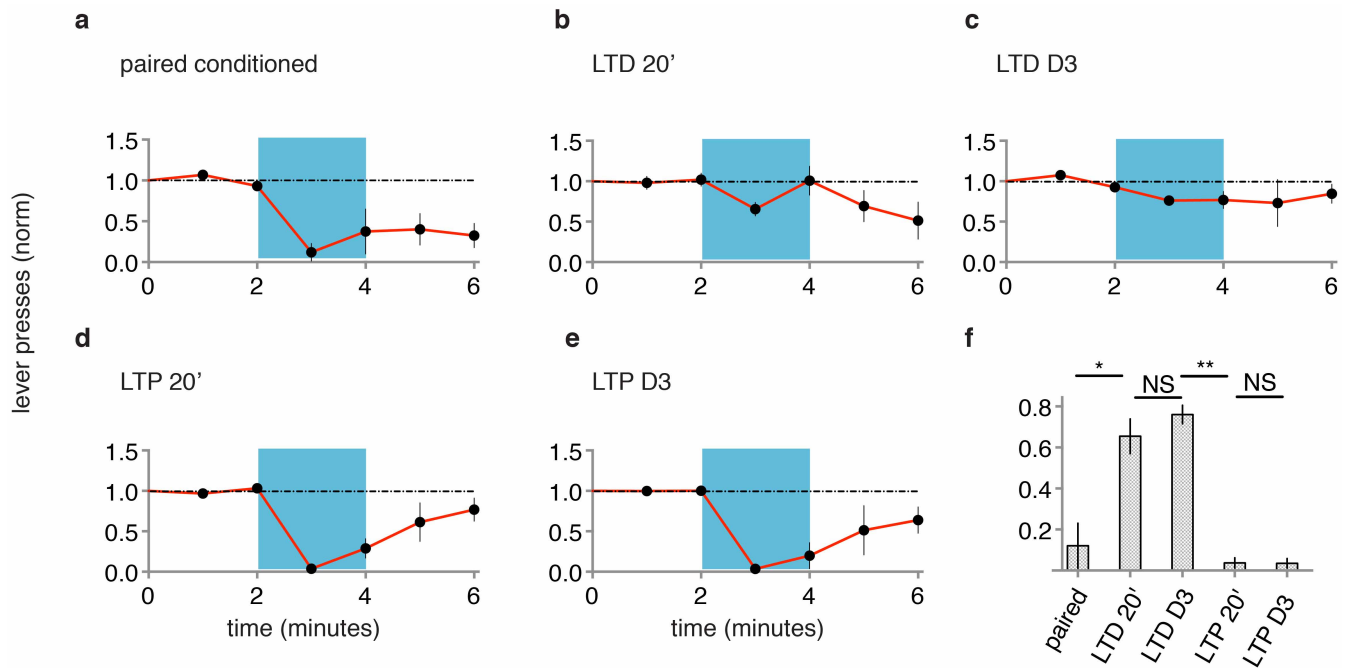
Extended Data Figure 6 | Systemic NMDA receptor blockade during conditioning blocks ODI-induced conditioned response. **a**, Animals ($n = 5$) were injected with MK801 (see Methods) and given optical CS paired with foot shock and subsequently tested one day later for CR. **b**, The same group of animals was then given optical CS paired with foot shock (in the absence of MK801) and subsequently tested one day later for CR. **c**, MK801 significantly blocked conditioning.



Extended Data Figure 7 | LTD and LTP remove and reactivate memory.

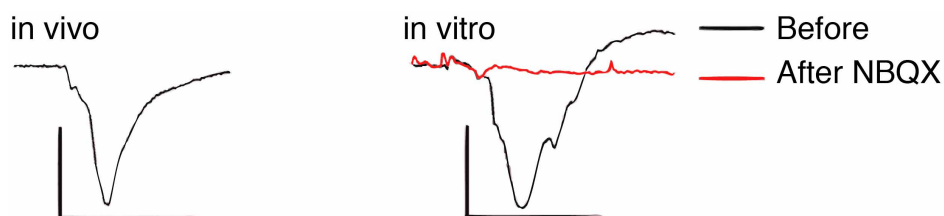
a–e, Data from an individual rat, measuring lever presses per minute before, during (blue) and after optical CS, one day after paired conditioning of optical CS and shock (**a**), one day after subsequent optical LTD protocol (**b**), one day

after subsequent optical LTP protocol (**c**), one day after subsequent second optical LTD protocol (**d**) and one day after subsequent second optical LTP protocol (**e**). **f**, Graph of lever presses during first minute into optical CS one day after delivery of indicated conditioning protocols.



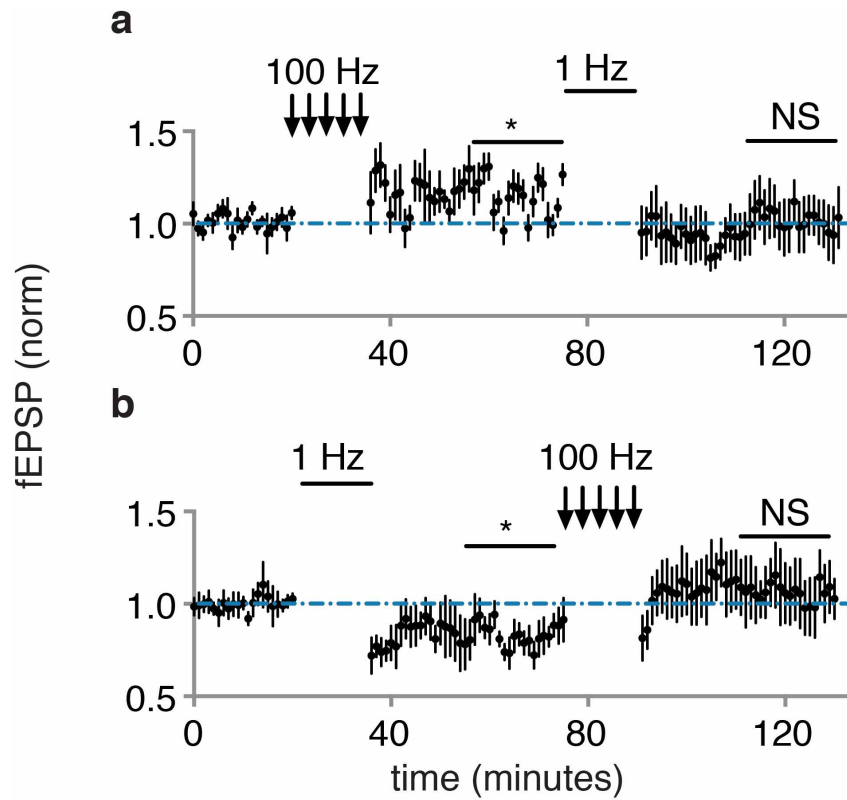
Extended Data Figure 8 | The effects of LTD and LTP are rapid and long-lasting. **a**, Animals ($n = 5$) were tested for CR one day following pairing of optical CS with shock. **b, c**, Within one hour of testing, animals received optical LTD protocol and were tested for CR 20 min (**b**) and three days (**c**) later.

d, e. Following day three, testing animals received optical LTP protocol and were tested for CR 20 min (**d**) and three days (**e**) later. **f.** Graph of normalized lever presses for the first minute of optical CS following indicated protocols.



Extended Data Figure 9 | Optically evoked *in vivo* and *in vitro* stimuli produce similar electrophysiological responses. Animals were injected *in vivo* with AAV-oChIEF-tdTomato in auditory regions 4 weeks before recordings. Left, *in vivo* electrophysiological response obtained from glass electrode placed in lateral amygdala and evoked by light pulse delivered

through fibre optic cable placed 500 μm above tip of glass electrode. Right, *in vitro* brain slice electrophysiological response obtained from glass electrode placed in lateral amygdala and evoked by light pulse delivered through fibre optic cable placed above the brain slice. Black trace is before and red trace after bath application of 10 μM NBQX. Scale bars, 1 mV, 10 ms.



Extended Data Figure 10 | LTD reverses LTP and LTP reverses LTD of *in vivo* optical responses in amygdala. **a**, Plot of baseline normalized fEPSP *in vivo* optically evoked responses ($n = 5$ from 5 rats) following optical LTP

(100 Hz) and optical LTD (1 Hz). **b**, Same as **a** for a separate group of recordings ($n = 5$) following optical LTD (1 Hz) and optical LTP (100 Hz). All comparisons to baseline period.

ABCB5 is a limbal stem cell gene required for corneal development and repair

Bruce R. Ksander^{1*}, Paraskevi E. Kolovou^{1*}, Brian J. Wilson^{2,3,4}, Karim R. Saab^{2,3}, Qin Guo^{2,3,4}, Jie Ma^{2,3}, Sean P. McGuire¹, Meredith S. Gregory¹, William J. B. Vincent¹, Victor L. Perez⁵, Fernando Cruz-Guilloty⁵, Winston W. Y. Kao⁶, Mindy K. Call⁶, Budd A. Tucker⁷, Qian Zhan⁸, George F. Murphy⁸, Kira L. Lathrop⁹, Clemens Alt¹⁰, Luke J. Mortensen¹⁰, Charles P. Lin¹⁰, James D. Zieske¹, Markus H. Frank^{2,3,11*} & Natasha Y. Frank^{2,4,11,12*}

Corneal epithelial homeostasis and regeneration are sustained by limbal stem cells (LSCs)^{1–3}, and LSC deficiency is a major cause of blindness worldwide⁴. Transplantation is often the only therapeutic option available to patients with LSC deficiency. However, while transplant success depends foremost on LSC frequency within grafts⁵, a gene allowing for prospective LSC enrichment has not been identified so far⁵. Here we show that ATP-binding cassette, sub-family B, member 5 (ABCB5)^{6,7} marks LSCs and is required for LSC maintenance, corneal development and repair. Furthermore, we demonstrate that prospectively isolated human or murine ABCB5-positive LSCs possess the exclusive capacity to fully restore the cornea upon grafting to LSC-deficient mice in xenogeneic or syngeneic transplantation models. ABCB5 is preferentially expressed on label-retaining LSCs² in mice and p63 α -positive LSCs⁸ in humans. Consistent with these findings, ABCB5-positive LSC frequency is reduced in LSC-deficient patients. *Abcb5* loss of function in *Abcb5* knockout mice causes depletion of quiescent LSCs due to enhanced proliferation and apoptosis, and results in defective corneal differentiation and wound healing. Our results from gene knockout studies, LSC tracing and transplantation models, as well as phenotypic and functional analyses of human biopsy specimens, provide converging lines of evidence that ABCB5 identifies mammalian LSCs. Identification and prospective isolation of molecularly defined LSCs with essential functions in corneal development and repair has important implications for the treatment of corneal disease, particularly corneal blindness due to LSC deficiency.

ABCB5, first identified as a marker of skin progenitor cells⁶ and melanoma stem cells^{7,9}, functions as a regulator of cellular differentiation⁶. On the basis of this function and its expression on stem cells in additional organ systems¹⁰, we hypothesized that ABCB5 might also identify slow-cycling, label-retaining LSCs in the eye. We performed bromodeoxyuridine (BrdU)-based ‘pulse-chase’ experiments (Extended Data Fig. 1a) in *Abcb5* wild-type mice, which revealed 8-week label-retaining cells only in the limbus, but not central cornea (Fig. 1a, b and Extended Data Fig. 1b). BrdU-retaining LSCs were located in basal limbal epithelium and demonstrated *Abcb5* co-expression (Fig. 1c, Extended Data Fig. 6c and Supplementary Videos 1 and 2). *Abcb5*⁺ cells (range 0.4–2.3%) were predominantly BrdU-positive (75.7 \pm 7.5%), in contrast to *Abcb5*[−] cells (3.3 \pm 2.3%, P < 0.001) (Fig. 1d). Similar to findings in mice (Figs 1c, 2d, e and Extended Data Fig. 3a, b), human ABCB5⁺ cells were also located in basal limbal epithelium (Fig. 1e). Moreover, they localized to the palisades of Vogt (Fig. 1e, Extended Data Fig. 1c–j and Supplementary Video 3). ABCB5⁺ limbal cells exclusively contained Δ Np63 α ⁺ human

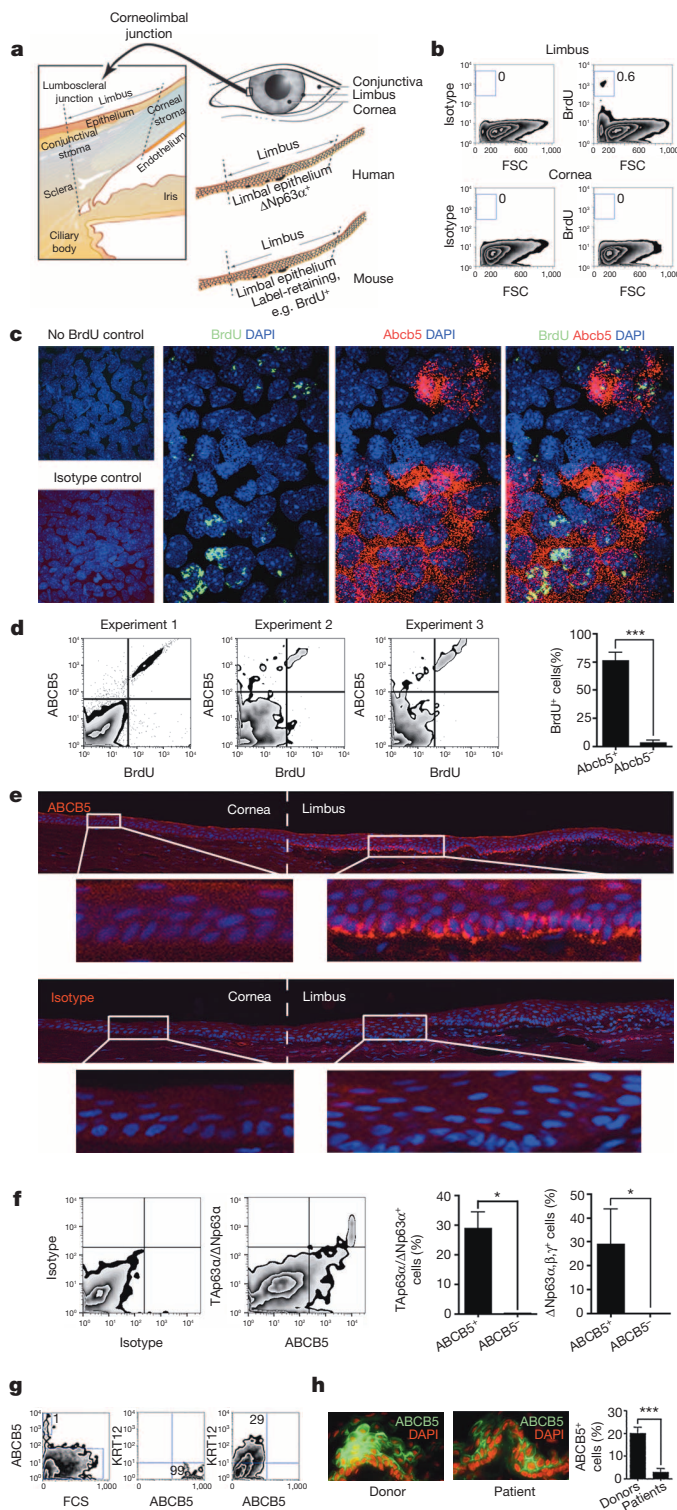
LSCs, determined using distinct Δ Np63 α antibodies (Δ Np63 α /TAp63 α epitope positivity in ABCB5⁺ versus ABCB5[−] cells: 28.9 \pm 5.7% versus 0.1 \pm 0.1%; Δ Np63 α , β , γ epitope positivity: 28.9 \pm 14.7% versus 0.1 \pm 0.1%; P < 0.05) (Fig. 1f) and did not express the differentiation marker keratin 12 (KRT12) (Fig. 1g). Moreover, limbal biopsies from LSC-deficient (LSCD) patients exhibited reduced ABCB5⁺ frequencies compared to controls (2.8 \pm 1.6% versus 20.0 \pm 2.6%, P < 0.001) (Fig. 1h and Extended Data Fig. 2). ABCB5 expression on label-retaining LSCs in mice and p63 α ⁺ LSCs in humans, along with reduced ABCB5⁺ frequency in clinical LSCD, showed that ABCB5 preferentially marks LSCs.

To investigate *Abcb5* function in corneal development and regeneration, we generated *Abcb5* knockout mice lacking exon 10 of the murine gene (GenBank accession number JQ655148), which encodes a functionally critical extracellular domain homologous to amino acids 493–508 of human ABCB5 (ref. 6) (GenBank accession number NM_178559) (Fig. 2a, b). Polymerase chain reaction (PCR) analysis confirmed deletion (Fig. 2c). *Abcb5* protein loss was demonstrated using an exon-10-encoded epitope-targeted monoclonal antibody (Fig. 2c), an amino-terminus-targeted antibody (Extended Data Fig. 3c), and a specific extracellular-loop-associated peptide-targeted human immunoglobulin (Ig)G1 monoclonal antibody (clone 3B9) (Fig. 2d and Extended Data Fig. 3a). Wild-type tissues only expressed *Abcb5* in the limbus but not the cornea (Fig. 2d and Extended Data Fig. 3a), consistent with findings in human tissues. Specificity of this binding pattern was demonstrated by RNA *in situ* hybridization (Fig. 2e and Extended Data Fig. 3b).

Abcb5 knockout mice were indistinguishable by physical examination from wild-type littermates through adulthood and their eyes contained all anterior and posterior segment components (Fig. 2f and Extended Data Fig. 3d). However, histological analysis of mutant versus wild-type corneas demonstrated developmental abnormalities characterized by decreased cellularity of the apical epithelial layer and disorganized basal and wing cell layers (Fig. 2f and Extended Data Fig. 3e). No inflammation was noted (Extended Data Fig. 3f). Reduced epithelial cell numbers in the central cornea but not limbus of knockout versus wild-type mice were confirmed by histological enumeration (cornea: 2,688 \pm 399 versus 4,427 \pm 346 cells, P < 0.05; limbus: 3,015 \pm 433 versus 3,629 \pm 94 cells, P = not significant) (Fig. 2g) and flow cytometry (Extended Data Fig. 3g). Corneas in knockout mice also exhibited epithelial tight junction defects (Fig. 2h) and increased fragility versus wild-type corneas (brush injury frequency 100% versus 33%, P < 0.001) (Extended Data Fig. 4a). Moreover, knockout versus wild-type mice showed reduced limbal and corneal Pax6 and corneal Krt12 expression (limbal Pax6:

¹Department of Ophthalmology, Schepens Eye Research Institute, Massachusetts Eye & Ear Infirmary and Harvard Medical School, Boston, Massachusetts 02114, USA. ²Transplant Research Program, Division of Nephrology, Boston Children's Hospital, Boston, Massachusetts 02115, USA. ³Department of Dermatology, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. ⁴Department of Medicine, VA Boston Healthcare System, Boston, Massachusetts 02130, USA. ⁵Bascorn Palmer Eye Institute and the Department of Ophthalmology, University of Miami Miller School of Medicine, Miami, Florida 33136, USA. ⁶Department of Ophthalmology, University of Cincinnati Medical Center, Cincinnati, Ohio 45229, USA. ⁷Stephen A Wynn Institute for Vision Research, Carver College of Medicine, Department of Ophthalmology and Visual Sciences, University of Iowa, Iowa City, Iowa 52242, USA. ⁸Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. ⁹Department of Ophthalmology, University of Pittsburgh School of Medicine & Department of Bioengineering, University of Pittsburgh Swanson School of Engineering, Pittsburgh, Pennsylvania 15213, USA. ¹⁰Center for Systems Biology and Wellman Center for Photomedicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA. ¹¹Harvard Stem Cell Institute, Harvard Medical School, Boston, Massachusetts 02138, USA. ¹²Division of Genetics, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA.

*These authors contributed equally to this work.



$0.3 \pm 0.3\%$ versus $18.0 \pm 4.6\%$; corneal Pax6: $8.3 \pm 4.6\%$ versus $42.0 \pm 7.6\%$; corneal Krt12: $6.5 \pm 6.5\%$ versus $47.7 \pm 8.2\%$; $P < 0.05$) (Fig. 2h and Extended Data Fig. 3h), also demonstrating an essential role of Abcb5 in corneal development. Additional ocular abnormalities involved the retina (Extended Data Fig. 5), where ABCB5 is also expressed¹¹.

Restoring the corneal epithelium after wounding is a hallmark function of LSCs. To determine whether wound healing requires Abcb5, knockout and wild-type mice received central corneal epithelial debridement injuries (Extended Data Fig. 4b–d). Wound closure rates were not significantly different (Extended Data Fig. 4e); however, knockout mice exhibited abnormal corneal restoration characterized by irregular

Figure 1 | ABCB5 marks LSCs. **a**, Cornea and LSC niche. **b**, BrdU detection (8-week chase). **c**, **d**, Immunofluorescence (×60 magnification) (**c**) and flow cytometric staining (gating based on control-staining) (**d**) for Abcb5/BrdU co-expression in mouse limbus. DAPI, 4',6-diamidino-2-phenylindole. Abcb5 and BrdU co-expression data in **d** represent analyses of $n = 4$ mice per group (mean \pm standard error of the mean (s.e.m.)). The experiment was performed three times. Data were analysed using the unpaired t -test, $P < 0.001$. **e**, ABCB5 positivity in human limbus (palisades of Vogt, ×20 magnification), with negativity in central cornea. **f**, ABCB5/p63α co-expression in human limbus. Quantitative analysis of ABCB5 monoclonal antibody and ΔNp63α/TAp63α epitope-binding antibody co-expression was performed using limbal epithelial cells from $n = 4$ eyes. The experiment was performed twice. Data were analysed using the unpaired t -test. Data are shown as mean \pm s.e.m., $P < 0.05$. The quantitative analysis of ABCB5 monoclonal antibody and ΔNp63α,β,γ epitope binding antibody co-expression was performed using limbal epithelial cells from $n = 4$ eyes. The experiment was performed twice. Data were analysed using the Mann–Whitney test. Data are shown as mean \pm s.e.m., $P < 0.05$. **g**, ABCB5/KRT12 co-expression. FSC, forward scatter. **h**, ABCB5 in LSCD patients (×20 magnification). Bar graph shows per cent ABCB5⁺ cells in control donors versus LSCD patients. Quantitative analysis of ABCB5 expression in $n = 2$ control and $n = 2$ LSCD specimens was performed using $n = 8$ sections per patient/control. All epithelial cells were counted in each section. A total of 2,031 and 2,051 epithelial cells were counted in patient 1 and donor 1, respectively. A total of 563 and 2,662 epithelial cells were counted in patient 2 and donor 2, respectively. Data were analysed using the unpaired t -test. Error bars indicate s.e.m. * $P < 0.05$, *** $P < 0.001$.

epithelium with reduced cell numbers (403 ± 30 versus 737 ± 28 , $P < 0.001$) (Fig. 2i and Extended Data Fig. 4f), increased cellular proliferation (Ki67 limbus: $54.0 \pm 5.0\%$ versus $0.3 \pm 0.2\%$, $P < 0.001$; cornea: $41.0 \pm 14.0\%$ versus $7.5 \pm 2.4\%$, $P < 0.05$) (Fig. 2j), and increased apoptosis (limbus: $41.2 \pm 12.8\%$ versus $1.0 \pm 0.5\%$; cornea: $49.0 \pm 10.0\%$ versus $0.4 \pm 0.3\%$; $P < 0.001$) (Fig. 2k and Extended Data Fig. 4g).

On the basis of our finding that Abcb5 preferentially identifies label-retaining and p63α⁺ LSCs, we hypothesized that Abcb5 was required for LSC maintenance. We therefore examined quiescent LSCs in knockout versus wild-type mice, using the BrdU-labelling approach (Extended Data Fig. 1a, b). After a 24 h chase, epithelial cells were equally labelled in knockout versus wild-type specimens ($5.7 \pm 0.9\%$ versus $7.2 \pm 1.0\%$, $P =$ not significant) (Extended Data Fig. 6b). In contrast, after an 8-week chase, label-retaining LSC frequency was reduced in knockout versus wild-type mice ($0.1 \pm 0.1\%$ versus $0.9 \pm 0.3\%$, $P < 0.05$) (Extended Data Fig. 6a). Histological examination of limbal epithelium confirmed selective loss of BrdU label-retaining cells in knockout versus wild-type mice (Extended Data Fig. 6c, d and Supplementary Video 2), demonstrating that abrogation of Abcb5 function induces LSC proliferation. Consistent with this result, Ki67 expression was enhanced in knockout versus wild-type tissues (limbus: $24.0 \pm 5.0\%$ versus $1.5 \pm 1.5\%$, $P < 0.001$; cornea: $53.0 \pm 16.0\%$ versus $11.0 \pm 2.1\%$, $P < 0.05$) (Extended Data Fig. 6e). Because cell cycle withdrawal is a prerequisite for LSC maintenance and hence normal differentiation, these results provided one explanation for the corneal differentiation defect in knockout mice. Moreover, impaired cell cycle withdrawal was associated with enhanced apoptosis, demonstrating a novel anti-apoptotic role of Abcb5. Consistent with this function, ABCB5 monoclonal antibody treatment of p63α-rich human limbal epithelial cells, using blocking concentrations¹², induced apoptosis in $30.9 \pm 2.9\%$ of cells versus controls ($P < 0.001$), commensurate with ABCB5 expression levels (Extended Data Fig. 6f, g). Moreover, ABCB5 blockade induced pro-apoptotic p53(S15) and p53(S392) and downregulated anti-apoptotic Bcl2 and Bcl-x (also known as Bcl2l1) (Extended Data Fig. 6h, i). In contrast, non-blocking monoclonal antibodies concentrations employed for cell sorting ($2 \mu\text{g ml}^{-1}$) maintained viability at $>90\%$.

Clinical studies in LSCD have shown that LSC frequency within grafts is critical for long-term transplant success⁵. To investigate whether ABCB5 represents a marker for prospective LSC enrichment, we examined the cornea-regenerative potential of transplanted murine or human

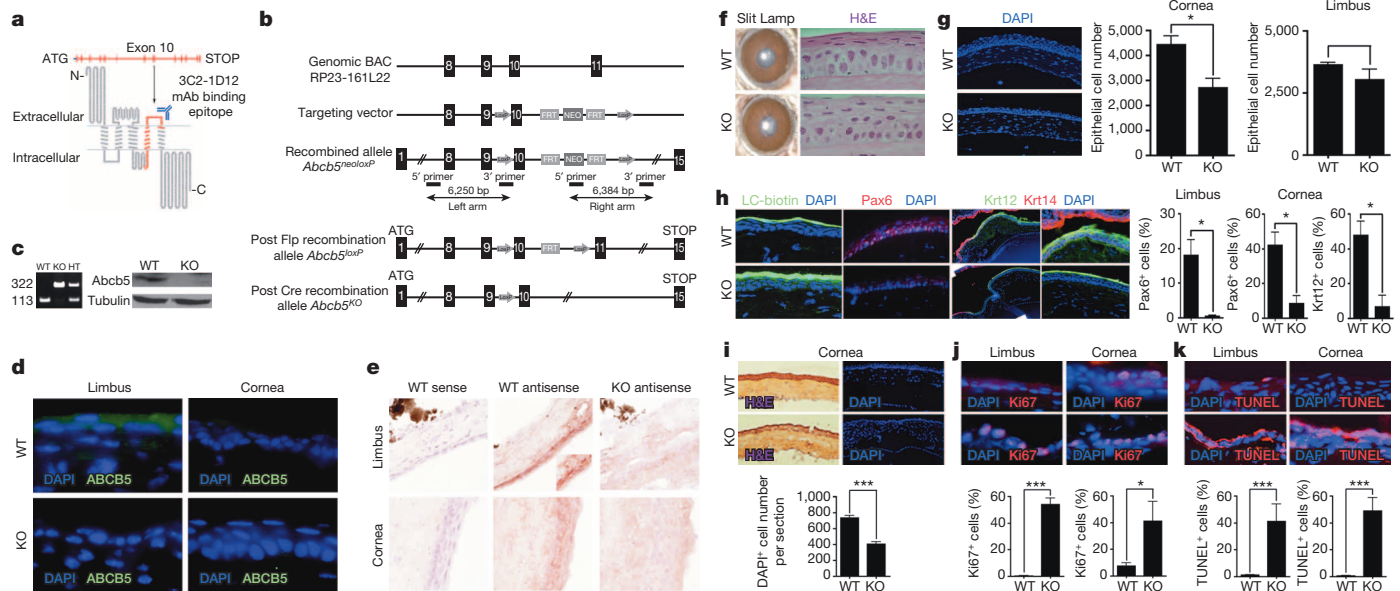


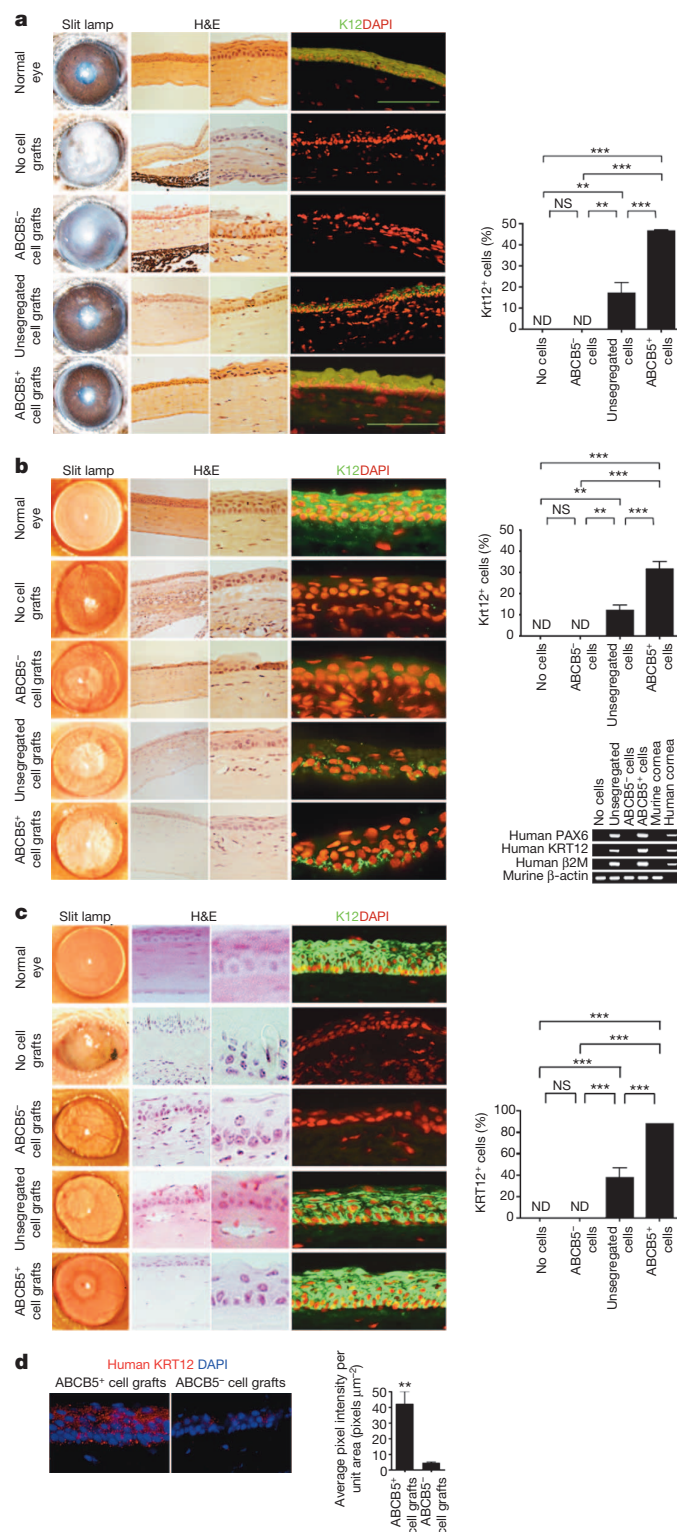
Figure 2 | ABCB5 regulates corneal development and repair. **a**, *Abcb5* locus and protein topology (transmembrane protein topology with a hidden Markov model (TMHMM), knockout deletion in red). mAb, monoclonal antibody. **b**, *Abcb5* knockout strategy. BAC, bacterial artificial chromosome. **c**, PCR and western blot in wild-type (WT) versus knockout (KO) mice. HT, heterozygous. **d**, **e**, *Abcb5* immunofluorescence staining (**d**) and *in situ* hybridization (**e**) of wild-type and knockout limbus and cornea ($\times 20$ magnification). **f-h**, Slit lamp and haematoxylin and eosin (H&E; $\times 40$ magnification) (**f**), cellularity (**g**) and LC-biotin diffusion and protein expression (**h**) analyses. **i**, Analysis of wild-type and knockout corneas following debridement wounding. **j**, **k**, Ki67 immunofluorescence (**j**) and TdT-mediated dUTP nick end labelling (TUNEL) (**k**) staining. **g-k**, $\times 20$ magnification. **g**, The numbers of viable epithelial cells in *Abcb5* knockout versus *Abcb5* wild-type murine central cornea were derived from the analysis of $n = 8$ mice per group (left bar graph). The experiment was performed four times. Data were analysed using the unpaired *t*-test. Data are shown as mean \pm s.e.m., $P < 0.05$. In the right bar graph, the numbers of viable epithelial cells in *Abcb5* knockout versus *Abcb5* wild-type murine limbus were derived from the analysis of $n = 6$ mice per group. The experiment was performed three times. Data were analysed using the unpaired *t*-test. Data are shown as mean \pm s.e.m., $P =$ not significant. **h**, Quantitative analyses of Pax6 expression in limbus and cornea of *Abcb5* knockout versus *Abcb5* wild-type mice were performed using $n = 3$ mice per group. The experiment was performed three times. Ten thousand cells per experiment were counted for each group. Data

were analysed using the unpaired *t*-test. Data are shown as mean \pm s.e.m., $P < 0.05$. Quantitative analyses of Krt12 expression in cornea were performed using $n = 2$ mice per group. The experiment was performed twice. Ten thousand cells per experiment were counted for each group. Data were analysed using the unpaired *t*-test. Data are shown as mean \pm s.e.m., $P < 0.05$. **i**, The numbers of DAPI⁺ cells per section in *Abcb5* wild-type versus *Abcb5* knockout mice (mean \pm s.e.m.) were derived from $n = 6$ mice per group. The experiment was performed twice. Within the standardized area (shown in Extended Data Fig. 4f), all corneal epithelial cells were counted in at least $n = 3$ consecutive composite cross-sections. Data were analysed using the unpaired *t*-test, $P < 0.001$. **j**, The percentages of Ki67⁺ epithelial cells in limbus and cornea of *Abcb5* knockout versus *Abcb5* wild-type mice (mean \pm s.e.m.) were determined using $n = 4$ mice per group. The experiment was performed twice. Within a standardized area, all limbal epithelial cells were counted in at least $n = 3$ consecutive cross-sections. Data were analysed using the unpaired *t*-test, $P < 0.001$ for limbus and $P < 0.05$ for cornea. **k**, The percentages of TUNEL⁺ epithelial cells in limbus or cornea of *Abcb5* knockout versus *Abcb5* wild-type mice (mean \pm s.e.m.) were determined using $n = 4$ mice per group. The experiment was performed twice. Within a standardized area, all limbal epithelial cells were counted in at least $n = 3$ consecutive cross-sections. Data were analysed using the unpaired *t*-test, $P < 0.001$ for limbus and $P < 0.001$ for cornea. Error bars indicate s.e.m. * $P < 0.05$, *** $P < 0.001$. NS, not significant.

unsegregated, ABCB5⁺ or ABCB5⁻ limbal cells in syngeneic or immunodeficient NSG mice with induced LSCD (Extended Data Fig. 7). Recipients of syngeneic murine ABCB5⁻ grafts or vehicle-only negative controls displayed opaque corneas, epithelial conjunctivalization and absence of differentiated Krt12⁺ cells (0%, in both cases) when analysed 5 weeks after transplantation, consistent with persistent LSCD (Fig. 3a and Extended Data Fig. 8a, b). Recipients of syngeneic unsegregated grafts displayed partial corneal restoration with differentiated Krt12⁺ cells in the central cornea (17%, enhanced versus ABCB5⁻ or vehicle-only treatment, $P < 0.01$), but exhibited persistence of LSCD-characteristic epithelial conjunctivalization (Fig. 3a and Extended Data Fig. 8a, b). In contrast, syngeneic ABCB5⁺ grafts resulted in the development of clear corneas with normal histology, gave rise to more differentiated Krt12⁺ cells (47%, increased versus unsegregated or ABCB5⁻ cell treatment or vehicle-only controls, $P < 0.001$) and prevented epithelial conjunctivalization (Fig. 3a and Extended Data Fig. 8a, b).

NSG recipients of human ABCB5⁺ grafts or vehicle-only controls also displayed epithelial conjunctivalization and an absence of differentiated KRT12⁺ cells (0%, in both cases) 5 weeks after transplantation (Fig. 3b and Extended Data Fig. 8c). NSG recipients of human unsegregated grafts

displayed partial corneal restoration with differentiated KRT12⁺ cells in the central cornea (12%, enhanced versus ABCB5⁻ or vehicle-only treatment, $P < 0.01$), but exhibited persistence of LSCD-characteristic epithelial conjunctivalization (Fig. 3b and Extended Data Fig. 8c). Strikingly, only human ABCB5⁺ grafts, negative for KRT12 before transplantation (Extended Data Fig. 7g), produced clear corneas with normal histology and high numbers of KRT12⁺ cells (31%, increased versus vehicle-only or versus ABCB5⁻ or unsegregated limbal cell treatment, $P < 0.001$) and an absence of LSCD-characteristic epithelial conjunctivalization (Fig. 3b and Extended Data Fig. 8c). To confirm that human donor cells caused corneal restoration, we assayed regenerated corneal tissue by PCR with reverse transcription (RT-PCR) for human-specific $\beta 2$ microglobulin ($\beta 2M$), an identifier of all human cells, and human-specific PAX6 and KRT12 as markers of corneal differentiation. Only corneal epithelium of recipients grafted with ABCB5⁺ or unsegregated human limbal cells contained human-specific $\beta 2M$, PAX6 and KRT12 transcripts, whereas vehicle-only-grafted control eyes that did not exhibit corneal restoration did not, confirming human specificity of the RT-PCR assay (Fig. 3b). Moreover, despite similar viability in ABCB5⁻ compared with unsegregated or ABCB5⁺ cell grafts (Extended Data Fig. 7e, f, h), ABCB5⁻ grafted



eyes were deficient in human-specific β2M, PAX6 or KRT12 expression (Fig. 3b), indicating that engraftment capacity was exclusively contained within the ABCB5⁺ cell population.

To further confirm LSC function of ABCB5⁺ limbal cells, we evaluated their capacity for long-term (>1 year) corneal restoration. LSCD-NSG recipients of human ABCB5⁺ grafts or vehicle-only controls, 13 months after transplantation, continued to display epithelial conjunctivalization, reduced epithelial thickness and increased stromal thickness with an absence of differentiated KRT12⁺ cells (0%, in both cases), consistent with persistent LSCD (Fig. 3c and Extended Data Figs 9, 10). NSG

Figure 3 | Regenerative role of ABCB5⁺ LSCs. **a, b**, Murine syngeneic grafts (**a**) and 5-week (**b**) or 13-month (**c**) human xenografts to LSCD-induced mice. Grafts contained either no donor cells (row 2), ABCB5⁻ cells (row 3), unsegregated cells (row 4), or ABCB5⁺ cells (row 5). **a–c**, Untreated C57BL/6 (**a**) and NSG (**b, c**) corneas (without LSCD) are shown in row 1. **b**, Bottom right, RT-PCR detection of human donor cells. **d**, Human KRT12⁺ cells (red) in recipient corneas of ABCB5⁺ human grafts at 13 months. Magnification: columns 2 and 4: ×20; column 3: ×40. Scale bars: 100 μm. The transplantation experiments shown in **a** and **b** were performed in *n* = 5 mice per group. The experiment was performed twice. For KRT12 expression analyses, all corneal epithelial cells within a standardized area (Extended Data Fig. 8b) were counted in at least *n* = 3 consecutive cross-sections from *n* = 4 replicate mice per group. Data were analysed using the one-way ANOVA and Bonferroni multiple comparisons tests. Data are shown as mean ± s.e.m. **c**, The transplantation experiments were performed in *n* = 5 mice per group. The experiment was performed twice. For KRT12 expression analyses, all corneal epithelial cells within a standardized area (Extended Data Fig. 9b) were counted in at least *n* = 3 consecutive cross-sections from *n* = 4 replicate mice per group. Data were analysed using the one-way ANOVA and Bonferroni multiple comparisons tests. Data are shown as mean ± s.e.m. **d**, Human KRT12 expression was analysed in *n* = 4 mice per group. Within a standardized area (Extended Data Fig. 9b), all corneal epithelial cells were counted in at least *n* = 3 consecutive cross-sections. Data were analysed using the unpaired *t*-test. Data are shown as mean ± s.e.m. Error bars indicate s.e.m. ***P* < 0.01, ****P* < 0.001. ND, not detected; NS, not significant.

recipients of human unsegregated grafts displayed partial corneal restoration with differentiated KRT12⁺ cells in the central cornea (37%, enhanced versus ABCB5⁻ or vehicle-only treatment, *P* < 0.001), but exhibited persistence of LSCD characterized by lower-than-normal epithelial thickness and higher-than-normal stromal thickness (Fig. 3c and Extended Data Figs 9, 10). In contrast, only purified human ABCB5⁺ grafts produced clear corneas with normal histology in recipient NSG mice, with the presence of a stratified epithelial layer containing high numbers of KRT12⁺ cells (88%, increased versus vehicle-only or versus ABCB5⁻ or unsegregated limbal cell treatment, *P* < 0.001) and the absence of LSCD-characteristic epithelial conjunctivalization, accompanied by restoration of normal epithelial and stromal thickness (Fig. 3c and Extended Data Figs 9, 10) and specific detection of human KRT12⁺ corneal cells (pixel intensity per unit area: 42.3 ± 7.7 versus 4.3 ± 0.7, respectively, *P* < 0.01) (Fig. 3d and Extended Data Fig. 9c).

Our findings that ABCB5⁺ cell frequency is reduced in LSCD, that ABCB5-positivity preferentially characterizes slow-cycling and p63α-positive populations, and that prospectively isolated ABCB5⁺ limbal cells are exclusively capable of reversing LSCD through long-term corneal regeneration show that ABCB5-positivity defines LSCs. This result is further supported by our demonstration that *Abcb5* knockout mice have impaired LSC-dependent corneal development and wound healing through deficient LSC maintenance due to deregulated anti-apoptotic signals. These results have several important implications. First, successful enrichment of human LSCs has the potential to decisively advance the field of LSCD therapy, because long-term clinical success depends on LSC frequency within grafts⁵ and because, thus far, no marker for prospective enrichment of bona fide LSCs defined by long-term corneal restorative capacity has been available¹³. Indeed, our study provides initial proof-of-principle for the hypothesis that prospective LSC enrichment within grafts can markedly enhance LSCD therapeutic success. What makes ABCB5 very useful and unique among LSC genes is its expression on the LSC surface, allowing for monoclonal-antibody-based LSC sorting strategies and enrichment as demonstrated in our study, in contrast to intracellularly expressed p63α and alternative candidate LSC-associated genes that have as of yet not been successfully employed for prospective isolation of human LSCs capable of *in vivo* LSCD reversal. This underscores the promise of ABCB5 as a marker for LSC isolation for clinical transplantation. Second, our study reveals a novel *in vivo* physiological function of ABCB5 in maintaining quiescent LSCs, through ABCB5-dependent regulation of apoptotic signalling pathways. This parallels the known anti-apoptotic function of the ABCB5

homologue ABCB1, shown to be mediated through cross-talk with Bcl-x¹⁴. Finally, our finding that ABCB5 regulates stem cell maintenance is highly relevant to the study of additional ABCB5-expressing normal stem cell populations in other tissues^{6,10} and of slow-cycling ABCB5⁺ cancer stem cells, in which this role might represent one mechanism of multidrug resistance to cell-cycle-specific agents^{10,11,15}. The herein described creation of a novel *Abcb5* gene knockout model represents a critical step towards such studies and to further dissection of *ABCB5* gene function in many relevant normal and cancerous tissues.

METHODS SUMMARY

Commercially available mouse strains and *Abcb5* knockout mice were maintained in accordance with the Institutional Guidelines of Boston Children's Hospital and the Schepens Eye Research Institute, Harvard Medical School. Human corneal tissues derived from consented donors according to Institutional Review Board (IRB)-approved protocols were obtained from Heartland Lions Eye Banks, the Bascom Palmer Eye Institute and the Carver College of Medicine.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 December 2012; accepted 30 April 2014.

Published online 2 July 2014.

- Davanger, M. & Evensen, A. Role of the pericorneal papillary structure in renewal of corneal epithelium. *Nature* **229**, 560–561 (1971).
- Cotsarelis, G., Cheng, S. Z., Dong, G., Sun, T. T. & Lavker, R. M. Existence of slow-cycling limbal epithelial basal cells that can be preferentially stimulated to proliferate: implications on epithelial stem cells. *Cell* **57**, 201–209 (1989).
- Majo, F., Roizat, A., Nicolas, M., Jaoude, G. A. & Barrandon, Y. Oligopotent stem cells are distributed throughout the mammalian ocular surface. *Nature* **456**, 250–254 (2008).
- Dua, H. S., Joseph, A., Shanmuganathan, V. A. & Jones, R. E. Stem cell differentiation and the effects of deficiency. *Eye (Lond.)* **17**, 877–885 (2003).
- Rama, P. *et al.* Limbal stem-cell therapy and long-term corneal regeneration. *N. Engl. J. Med.* **363**, 147–155 (2010).
- Frank, N. Y. *et al.* Regulation of progenitor cell fusion by ABCB5 P-glycoprotein, a novel human ATP-binding cassette transporter. *J. Biol. Chem.* **278**, 47156–47165 (2003).
- Schatton, T. *et al.* Identification of cells initiating human melanomas. *Nature* **451**, 345–349 (2008).
- Pellegrini, G. *et al.* p63 identifies keratinocyte stem cells. *Proc. Natl Acad. Sci. USA* **98**, 3156–3161 (2001).
- Luo, Y. *et al.* Side population cells from human melanoma tumors reveal diverse mechanisms for chemoresistance. *J. Invest. Dermatol.* **132**, 2440–2450 (2012).
- Wilson, B. J. *et al.* ABCB5 identifies a therapy-refractory tumor cell population in colorectal cancer patients. *Cancer Res.* **71**, 5307–5316 (2011).
- Frank, N. Y. *et al.* ABCB5-mediated doxorubicin transport and chemoresistance in human malignant melanoma. *Cancer Res.* **65**, 4320–4333 (2005).
- Lin, J. Y. *et al.* Genetically determined ABCB5 functionality correlates with pigmentation phenotype and melanoma risk. *Biochem. Biophys. Res. Commun.* **436**, 536–542 (2013).
- Notara, M. & Daniels, J. T. Biological principals and clinical potentials of limbal epithelial stem cells. *Cell Tissue Res.* **331**, 135–143 (2008).
- Rocco, A. *et al.* MDR1-P-glycoprotein behaves as an oncofetal protein that promotes cell survival in gastric cancer cells. *Lab. Invest.* **92**, 1407–1418 (2012).
- Cheung, S. T., Cheung, P. F., Cheng, C. K., Wong, N. C. & Fan, S. T. Granulin-epithelin precursor and ATP-dependent binding cassette (ABCB5) regulate liver cancer cell chemoresistance. *Gastroenterology* **140**, 344–355 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank R. Maas for critical reading of the manuscript. We also thank P. Mallen for assistance with graphic illustrations; the Heartland Lions Eye Bank for providing clinical specimens; G. Berg for assistance with mouse colony maintenance; D. Dombkowski, F. Pfeffer and R. Huang for their assistance with cell sorting and viability studies. This work was supported by National Institutes of Health (NIH)/National Institutes of Neurological Disorders and Stroke grant K08NS051349, VA BLR&D 1101BX000516 and VA RR&D 1101RX000989 Merit Review Awards, and a Harvard Stem Cell Institute grant to N.Y.F., NIH/National Cancer Institute grants R01CA113796, R01CA158467 and R01CA138231 to M.H.F., Department of Defense grant PR0332453 to B.R.K., NIH R01EY018624 and P30EY014801 grants to V.L.P., NIH grant R01EY021768 to W.W.K., NIH New Innovator Award DP2OD007483 and a Corley Research Foundation grant to B.A.T., NIH R01CA138231 to G.F.M., Western Pennsylvania Medical Eye Bank Core Grant for Vision Research (EY08098) to K.L.L., and NIH grants R01 EB017274, U01HL100402 and P41EB015903 to C.P.L. F.C.G. is a Howard Hughes Medical Institute Fellow of the Life Sciences Research Foundation.

Author Contributions N.Y.F., M.H.F. and B.R.K. designed the study. P.E.K., B.R.K., N.Y.F., M.H.F., B.J.W., K.R.S., Q.G., J.M., S.P.M., M.S.G., W.J.B.V., Q.Z., K.L.L., C.P.L., C.A. and L.J.M. performed experiments. V.L.P., F.C.G. and B.A.T. provided reagents and specimens. W.W.Y.K. and M.K.C. provided technical assistance. P.E.K., B.J.W., Q.G., J.M., S.P.M., M.S.G., W.J.B.V., J.D.Z., G.F.M., B.R.K., M.H.F. and N.Y.F. analysed the data. N.Y.F., M.H.F. and B.R.K. wrote the manuscript. M.H.F. and N.Y.F. are co-senior investigators.

Author Information The murine *Abcb5* messenger RNA sequence has been deposited in GenBank under accession number JQ655148. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.Y.F. (nfrank@partners.org) or M.H.F. (markus.frank@childrens.harvard.edu).

METHODS

Animals. Male and female C57BL/6J, NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ (NSG), B6;SJL-Tg(ACTFLPe)9205Dym/J and B6.FVB-Tg(EIIa-cre)C5379Lmgd/J mice were purchased from Jackson Laboratory. *Abcb5* knockout mice were generated as described later. All animals were maintained in accordance with the Institutional Guidelines of the Boston Children's Hospital and the Schepens Eye Research Institute, Harvard Medical School. Four-to-twelve-week-old mice were used for the experiments.

Generation of a germline *Abcb5* knockout mouse. We generated a conditional knockout construct by inserting two *loxP* sites flanking murine *Abcb5* exon 10 (GenBank accession number JQ655148) (Fig. 2a, b). A targeting construct was generated by recombineering¹⁶. In brief, a neomycin resistance cassette flanked by two *loxP* sites (based on plasmid pL-452) was inserted into the BAC clone RP23-161L22 458 base pairs upstream of exon 10 of the murine *Abcb5* gene (GenBank accession number JQ655148). The targeted region of the BAC clone was retrieved by gap repair into the pL-253 plasmid. The retrieved plasmid contained 6,006 bp upstream of exon 10 (not including the inserted neo cassette) and 6,384 bp downstream of exon 10. The neomycin resistance cassette was excised by arabinose induction of Cre recombinase to leave a single *loxP* site upstream of exon 10. A neomycin resistance cassette flanked by two FRT sites and one *loxP* site (based on plasmid pL-451) was inserted 460 bp downstream of exon 10 to complete the targeting construct. The targeting plasmid was verified by DNA sequencing and restriction mapping. The linearized plasmid was transfected into TC1 (129S6/SvEvTac derived) embryonic stem (ES) cells and selected in G418 (Sigma-Aldrich) and Fialuridine (Moravak Biochemicals). Resistant colonies were expanded and screened by long-range PCR to identify targeted clones. The left arm was amplified with 5'-GTTGAGGGGAGCAGCCAGAGCAAGGTGAGAAAGGTG-3' and 5'-TTAAGGGTTATTGAATATGATCGGAATTGGGCTGCAGGAATT-3' primers yielding a 6,250 bp PCR product (Fig. 2b). The right arm was amplified with 5'-TG GGGCAGGACAGCAAGGGGAGGAT-3' and 5'-CTGGTCCCTCTCTCTGTG ATCTACACAGGCC-3' primers yielding a 6,384 bp PCR product (Fig. 2b). Two *Abcb5*-targeted ES clones were identified. These clones were expanded and injected into C57BL/6 blastocysts that were then transferred to the uterus of pseudopregnant females. High-percentage chimaeric male mice (*Abcb5*^{neo-*loxP*/wt}) were bred into a C57BL/6 background to obtain germline transmission. Germline transmission of the *Abcb5*^{neo-*loxP*} allele was confirmed by PCR analysis of genomic DNA using 5'-GGAAGACAATAGCAGGCATGCTGGG-3', 5'-GGCTGGGGCAACTGAAAA GTAGCAT-3', and 5'-TTTCAGCTTCAGTTTATCACAATGTGGGTT-3' primers designed to amplify the 385 bp targeted allele and the 284 bp wild-type allele. Heterozygous *Abcb5*^{neo-*loxP*} mice were then intercrossed with hACTB-FLPe transgenic mice¹⁷ to remove the neomycin resistance cassette. PCR analysis of genomic DNA was performed to confirm removal of the neomycin resistance cassette in the genome of *Abcb5*^{loxP/wt} mice using 5'-ACTTGGTGCGGTGACTCTGAATTTT GC-3' and 5'-TAGCAACATTTCTGGCATTCTTAGGCTG-3' primers designed to amplify a 494 bp neomycin resistance cassette-deleted allele and a 390 bp wild-type allele. To determine the outcome of a complete loss of ABCB5 function, exon 10 of the murine *Abcb5* gene was deleted by breeding *Abcb5*^{loxP} mice with EIIa-Cre mice, which express Cre recombinase at the zygote stage^{18,19} (Fig. 2b). Deletion of the genomic region between the two *loxP* sites was confirmed by PCR analysis of genomic DNA using 5'-GGCTGGGGCAACTGAAAAAGTAGCAT-3', 5'-GCAAA TGTGTACTCTGCGCTTATTAATG-3' and 5'-TGGTGCAGACTACAGACG TCAGTGG-3' primers designed to amplify a 322 bp cre-deleted allele (null) and a 113 bp wild-type (WT) allele (Fig. 2c). Heterozygous *Abcb5*^{null/wt} mice with the germline deletion of exon 10 were intercrossed to produce homozygous *Abcb5*^{null/null} (*Abcb5* knockout) mutants. Mice were maintained on a 129S6/SvEvTac/C57BL/6 mixed genetic background and littermates were used as controls for experimental analyses.

Western blot analysis and human apoptosis array. *Abcb5* wild-type and *Abcb5* knockout cell lysates were immunoblotted using monoclonal ABCB5 antibody 3C2-1D12 (ref. 6) (5.5 µg ml⁻¹), a rabbit polyclonal N-terminus-targeted ABCB5 antibody (1:100 dilution) (Abgent), or an α-tubulin rabbit polyclonal antibody (1:5,000 dilution) (Abcam). A human apoptosis proteome profiler antibody array (R+D Systems) was used according to the manufacturer's instructions, using 400 µg human limbal epithelial cell lysates prepared from cells treated for 48 h with either blocking concentrations¹² of anti-ABCB5 monoclonal antibody clone 3C2-1D12 (ref. 6) or equivalent concentrations of isotype control monoclonal antibody (clone MOPC31C, Sigma). The pixel densities of array spots were quantified using ImageJ software.

RT-PCR. For detection of human-specific gene transcripts, total RNA was isolated from transplanted murine eyes and non-injured murine or human control corneas using the RNeasy Plus isolation kit (Qiagen) and then transcribed using the High Fidelity RT kit (Applied Biosystems). PCR was performed using Taq 2X Master Mix (New England Biolabs) and the following gene-specific primers.

Human β2-microglobulin (B2M, NM_004048): forward 5'-GTGTCTGGGTT TCATCCATC-3', reverse 5'-AATGCGGCATCTTCAACCTC-3'; human paired box 6 (PAX6, NM_000280.3): forward 5'-CAGCGCTCTGCCGCCTAT-3', reverse 5'-CATGACCAACACAGATCAAACATCC-3'; human keratin 12 (KRT12, NM_000223.3): forward 5'-GAAGCCGAGGGCGATTACTG-3', reverse 5'-GTGCTTG TGATTGGAGTCTGTAC-3'; and murine β-actin (Actb, NM_007393): forward 5'-TCCTAGCACCATGAAGATC-3', reverse 5'-AAACGCAGCTCAGTAACA G-3'.

RNA *in situ* hybridization. *Abcb5* RNA probes were prepared as follows. PCR-derived RNA probe templates were synthesized by introducing the T7 promoter into the antisense strand and the SP6 promoter into the sense strand. The primer pair (5'-TAATACGACTCACTATAGGACCATATGCAATGGCGGTAAAG-3' and 5'-GATTAGGTGACACTATAGAGACACTTCAAGACTCAACACAG-3') was used to generate the DNA template for antisense and sense RNA probes spanning 144 bp of murine *Abcb5* complementary DNA containing exon 10 (GenBank accession number JQ655148). RNA probe labelling with digoxigenin (DIG) and *in situ* hybridizations were carried out as described previously²⁰.

BrdU pulse and chase experiments. Four-week-old *Abcb5* knockout mice and their wild-type littermates were subjected to daily intraperitoneal injections of 50 mg kg⁻¹ BrdU (BD Pharmingen) for 9 consecutive days (Extended Data Fig. 1a). Limbal and central corneal epithelial cells were harvested from *Abcb5* wild-type and *Abcb5* knockout mice either 24 h or 8 weeks after receiving the last BrdU injection. Limbal and central corneal epithelial cells were also harvested from age-matched untreated *Abcb5* wild-type and *Abcb5* knockout mice for use as experimental controls. Flow cytometry and immunohistochemistry staining were used to determine the frequency of BrdU-positive and BrdU-negative cells within epithelia of the limbus and central cornea. The threshold for BrdU positivity in the pulse chase experiments was determined using the background levels obtained from anti-BrdU antibody-stained limbal or corneal epithelial cells from either wild-type or *Abcb5* knockout mice that did not receive any prior BrdU injections. These thresholds were used to establish the BrdU-positive gates shown in Extended Data Fig. 6a. An example of the controls used to set gates for BrdU positivity is shown in Extended Data Fig. 1b.

Human and murine corneal cell isolation. Cadaveric human corneal scleral tissues derived from consented donors according to Institutional Review Board (IRB)-approved protocols were obtained from Heartland Lions Eye Banks, Bascom Palmer Eye Institute, and Carver College of Medicine. After removal of the scleral rim, iris and trabecular meshwork, the limbus and central cornea were dissected under a microscope. Limbal and central corneal tissues were subsequently incubated with 2.4 units ml⁻¹ Dispase II (Roche Diagnostics) at 37 °C for 1 h, followed by incubation with 0.5 M EDTA (Invitrogen) at 37 °C to recover the epithelial cells^{21,22}. Murine limbal and corneal epithelial cells were obtained from *Abcb5* knockout and *Abcb5* wild-type mice. Immediately after euthanasia by CO₂ narcosis and subsequent eye enucleation, limbal and central corneal tissues were removed with microscissors under a dissecting microscope, placed in low Ca²⁺ keratinocyte serum free medium (KSFM, Invitrogen) and centrifuged for 5 min at 250g at 4 °C. After removal of the supernatant, tissue pellets were digested in 0.5% trypsin solution (Lonza)²³. For transplantation experiments, ABCB5⁺ and ABCB5⁻ limbal epithelial cells were isolated by FACS using ABCB5 monoclonal antibody labelling¹¹. In brief, either human or murine limbal epithelial cells were labelled with primary ABCB5 monoclonal antibody (20 µg µl⁻¹) for 30 min at 4 °C, washed to remove excess antibody, followed by a 30 min incubation with secondary anti-mouse FITC-conjugated IgG. The ABCB5⁺ and ABCB5⁻ sorting gates were established on a Modified Digital Vantage cell sorter (Becton Dickinson and MGH Pathology Flow Cytometry Core, Simches Research Building) as shown in Extended Data Fig. 7. Only viable cells were selected for sorting by excluding all DAPI⁺ cells (1 µg ml⁻¹ DAPI, Sigma-Aldrich, added immediately before sorting) as identified using a 70 mW UV laser for excitation. The purity and viability of ABCB5⁺ and ABCB5⁻ sorted cells were established in representative post-sort analyses in which samples were re-analysed (Extended Data Fig. 7). ABCB5⁺ cell purification resulted in a 255-fold increase for murine ABCB5⁺ limbal cells (0.37% positivity before and 51% positivity after sorting; Extended Data Fig. 7h) and a 292-fold increase for human ABCB5⁺ limbal cells (0.03% positivity before and 59% positivity after sorting; Extended Data Fig. 7h). ABCB5⁻ cell enrichment resulted in complete absence of ABCB5⁺ cells in both mouse and human samples (Extended Data Fig. 7h). Primary culture-expanded p63α-rich limbal epithelial cells for use in ABCB5 inhibition studies were purchased from Invitrogen (catalogue no. C-018-5C).

Flow cytometric analysis. Dual-colour flow cytometry was used to determine whether human ABCB5⁺ limbal epithelial cells co-expressed p63α or KRT12 and whether murine *Abcb5*⁺ limbal epithelial cells co-expressed Pax6 and Krt12, and was performed as described previously¹¹. For human and murine ABCB5 and KRT12 co-expression analysis, cells were first incubated with mouse anti-ABCB5 monoclonal antibody, counterstained with goat anti-mouse FITC IgG, followed by incubation

with goat polyclonal anti-KRT12 antibody and counterstaining with Dylight 649 donkey anti-goat IgG. For human ABCB5 and p63 α co-expression and murine Abcb5 and Pax6 co-expression analysis, cells were incubated with mouse anti-ABCB5 monoclonal antibody, counterstained with goat anti-mouse FITC IgG, permeabilized in BD Cytofix/Cytoperm Buffer (BD Biosciences), stained with either p63 α or Pax6 antibodies, and counterstained with goat anti-rabbit Alexa 647 IgG. Washing steps with staining buffer or BD Perm/Wash Buffer (BD Biosciences) were performed between each step. Dual-colour flow cytometry was performed by acquisition of fluorescence emission at the FL1 (FITC) and FL4 (Alexa 647 and/or Dylight 649) spectra on a Becton Dickinson FACScan (Becton Dickinson), as described¹¹. Murine Abcb5 and BrdU co-expression analysis was performed using the FITC BrdU Flow Kit (BD Biosciences), according to the manufacturer's instructions. Statistical differences between expression levels of the above-listed markers by ABCB5⁺ and ABCB5⁻ cells were determined using the unpaired *t*-test. A two-sided *P* value of *P* < 0.05 was considered significant. For determination of the epithelial cell numbers in the central cornea and limbus of *Abcb5* wild-type and *Abcb5* knockout mice, dissociated single cell suspensions were stained with DAPI (1 μ g ml⁻¹ DAPI, Sigma-Aldrich) and analysed by flow cytometry using forward-scattered light (FSC) versus side-scattered light (SSC) to identify viable cells. Statistical differences between *Abcb5* wild-type and *Abcb5* knockout mice were determined using the unpaired *t*-test. A two-sided *P* value of *P* < 0.05 was considered significant.

Histopathology and immunohistochemical staining. To recover intact mouse ocular tissue, the whole decapitated mouse head was fixed in 4% PFA overnight, then eyes were enucleated with the lids attached, incubated in 30% sucrose in 1 \times PBS overnight at 4 °C, embedded in Tissue-Tek OCT compound (Sakura Finetek USA) and snap-frozen. Representative cryostat sections from each tissue block were stained with H&E. For immunofluorescence staining, cryostat sections (10 μ m) were fixed in cold methanol for 10 min, blocked in 10% secondary serum plus 2% BSA in 1 \times PBS for 1 h, incubated with the primary antibody (or isotype control), followed by the appropriate secondary antibody for 1 h at room temperature. Following several washes, the slides were then coverslipped in hard-set mounting media with DAPI. Composite corneal photographs were assembled using Photoshop (Adobe) to overlay and match sequential images. Stitching was done by reducing the added photograph to 50% transparency, matching images, and returning the composite photograph to 0% transparency. The average number of epithelial cells per cornea (Fig. 2g) was determined by counting the number of DAPI-positive cells within the area defined by a 2 mm trephine in a composite photograph of a complete corneal section. At least three composite corneal sections were analysed per mouse, and five mice were analysed per group in four replicate experiments. The percentages of epithelial cells expressing Ki67 (Fig. 2j and Extended Data Fig. 6e), TUNEL (Fig. 2k) and Krt12 (Fig. 3a, b) were determined by counting the number of positive cells among the total number of DAPI-positive corneal epithelial cells using the techniques described earlier. Comparisons between the *Abcb5* wild-type and *Abcb5* knockout mice were performed using the unpaired *t*-test. The results of transplantation experiments were compared using the analysis of variance (ANOVA) test. Differences with *P* < 0.05 were considered statistically significant. For preparation of murine cornea whole mounts used for BrdU and Abcb5 immunostaining, whole mouse eyes with lids attached were enucleated, rinsed in PBS and immediately fixed in 4% PFA overnight at room temperature. Fixed eyes were washed in PBS, the globe was bisected under a dissecting microscope in the nasal/temporal axis and the lids were removed. The posterior half of the eye was removed, leaving the cornea, limbus and part of the sclera intact. Relief cuts (7–10) were made from the corneal side and also from the scleral side to allow the limbus to lie as flat as possible. Throughout the dissection care was taken to handle the cornea and limbus as little as possible in order to retain anatomical integrity. To stain whole mount tissue for BrdU, tissue was placed on a rocker in 2 N HCl for 30 min, followed by 20 min in 0.1 M sodium borate. Two rinses in PBS were followed by 1 h in blocking buffer (25 ml 250 mg BSA, 2.5 ml 10 \times PBS, 1.5 ml goat serum, 75 μ l Triton X-100, distilled H₂O). Two more PBS rinses were followed by overnight incubation with the following primary antibodies: rabbit anti-ABCB5 antibody (NBP1-77687, Novus) and biotinylated mouse anti-BrdU antibody (51-75512x, BD Pharmingen). Next, after two PBS rinses, tissue was incubated in blocking buffer for 15 min, washed twice in PBS, and incubated with secondary antibodies for 1 h (Alexa Fluor546 goat anti-rabbit and APC Streptavidin). Two more PBS rinses were followed by 5 min of DAPI incubation and two further PBS rinses. Tissue was then mounted on slides (Immu-Mount; Thermo-electron Incorporated) with spacers to preserve the morphology (Avery laser labels, 5,262 with holes punched for tissue) and subsequently coverslipped. Confocal microscopy was conducted on an inverted laser scanning confocal microscope system (FV1000, Olympus) with an automated stage (Prior Scientific) with $\times 20$ (RI 0.85) and $\times 60$ (RI 1.42) objectives.

Detection of palisades of Vogt and ABCB5 confocal microscopy. Optical coherence tomography (OCT) imaging of human corneal rims was performed with a prototype system as described previously by Lathrop *et al.*²⁴. In brief, globes were

punctured in several areas before the overnight fixation in 4% PFA. The following day, the tissue was washed and kept in PBS until ready to be embedded and sectioned. Prior to sectioning, the globes were scanned with a modified Biotigen spectral-domain OCT system (Biotigen; SuperLum) in order to define limbal regions containing defined palisade structures (Extended Data Fig. 1c–j). These palisade-containing regions were marked with a surgical pen (Extended Data Fig. 1c). Corneal and limbal tissues were then dissected from the marked areas (1 cm wide, extending 1 cm into the cornea and 1 cm into the sclera, see Extended Data Fig. 1i, j) and sunk in a sucrose solution (15%, then 30%) in order to preserve the tissue structure during embedding. Tissue sections were placed in moulds surrounded by cooled (4 °C) OCT compound (Tissue-Tek, Sakura Finetek USA) and frozen in liquid nitrogen. Thirty-micrometre palisade-containing sections were cut on a cryostat refrigerated microtome and placed on slides (Fisher Superfrost, Thermo Fisher Scientific). Prior to immunolabelling, slides were washed in distilled water 3 \times 15 min and once in NaBH₄ (1 mg ml⁻¹) for 10 min, and permeabilized with blocking buffer (1 \times PBS, 1% BSA, 0.3% Triton X-100, 6% donkey serum) for 1 h at room temperature. Slides were subsequently incubated in the dark at room temperature for 1 h with anti-ABCB5 monoclonal antibody (clone 3C2-1D12) diluted in dilution buffer (1 \times PBS, 1% BSA) and washed 3 \times 5 min each in blocking buffer. Alexa Fluor546 donkey anti-mouse antibody dissolved in dilution buffer was applied for 1 h at room temperature. Subsequently, three washes of blocking buffer were applied for 5 min each, followed by nuclear staining, washing 2 \times 5 min in 1 \times PBS and mounting with Immu-Mount (Thermo Electron Corporation). Slides were examined by confocal microscopy conducted on an inverted laser scanning confocal microscope system (FV1000, Olympus) with an automated stage (Prior Scientific) using $\times 20$ oil (RI 0.85) and $\times 60$ (RI 1.42) objectives. The three-dimensional display of reconstructed stacks is available as Supplementary Video 3.

Antibodies. The following primary antibodies were used in flow cytometry experiments. Rabbit polyclonal anti-p63 α antibody (H-129, sc-8344, Santa Cruz), rabbit polyclonal anti-p40 (DNp63) antibody (ABX-144A, Imgenex), mouse monoclonal anti-ABCB5 antibody (clone 3C2-2D12)⁶, goat polyclonal anti-cytokeratin 12 antibody (L15, sc-17101, Santa Cruz), rabbit polyclonal anti-Pax6 antibody (PRB-278P, Covance), rabbit polyclonal IgG isotype control antibody (ab27478, Abcam), mouse IgG1 κ isotype control antibody (clone X40, BD Biosciences), and goat IgG isotype control antibody (sc-3887, Santa Cruz). The secondary antibodies were goat anti-mouse FITC (F2012, Sigma-Aldrich), Alexa 647 goat anti-rabbit IgG (A21244, Invitrogen) and Dylight 649 donkey anti-goat (Jackson ImmunoResearch). For histopathology and immunohistochemical analyses, the following antibodies were used. Mouse monoclonal anti-ABCB5 (clone 3C2-1D12)⁶ and rabbit antibody against p63 α at 1:75 dilution (H-129, sc8344, Santa Cruz) followed by the appropriate secondary antibodies obtained from Jackson ImmunoResearch: FITC-donkey anti-rabbit (711-095-152) at 1:75 dilution or Alexa Fluor594 goat anti-mouse (115-515-062) at 1:250 dilution. In all cases, the isotype-matched antibodies rabbit IgG (550875, BD Pharmingen) and mouse IgG1 κ isotype control antibody (clone X40, BD Biosciences) served as controls. Further antibodies used for histopathology and immunohistochemical analyses were as follows. Rabbit anti-ABCB5 antibody at 1:250 dilution (NBP1-50547, Novus), human anti-ABCB5 extracellular loop-associated peptide monoclonal antibody (clone 3B9, 5 μ g ml⁻¹), rabbit polyclonal anti-Pax6 at 1:300 dilution (PRB-278P, Covance), goat anti-cytokeratin 12 (L15) at 1:50 dilution (sc17101, Santa Cruz), rabbit anti-human cytokeratin 12 (H-60, sc-25722), rabbit anti-cytokeratin 14 (AF64) at 1:1,000 dilution (PRB-155P, Covance), rabbit anti-Ki67 at 1:200 dilution (ab66155, Abcam), biotinylated BrdU antibody (51-75512, Pharmingen), followed by the appropriate secondary antibodies obtained from Jackson ImmunoResearch: donkey anti-goat Alexa Fluor488 at 1:250 dilution (705-545-003), donkey anti-rabbit Alexa Fluor594 at 1:20 dilution (711-585-152), goat anti-rabbit DyLight 549 at 1:250 dilution (111-504-144), or Cy3-donkey anti-rabbit at 1:250 dilution (711-165-152). Appropriate isotype-matched antibodies (rabbit IgG (550875, BD Pharmingen) and goat IgG (sc2028, Santa Cruz)) served as negative controls. Additional primary antibodies included rat anti-mouse Ly-6G 1:100 dilution (550291, BD Biosciences), rat anti-mouse F4/80 1:200 dilution (RM2900, Abcam), rat anti-mouse CD45R 1:25 dilution (550286, BD Biosciences), rabbit anti-CD3 1:100 dilution (5690, Abcam), rat IgG isotype control (559073, BD Pharmingen) and rabbit IgG isotype control (NB810-56910, Novus). Secondary antibodies were Alexa Fluor555 goat anti-rat at 1:250 dilution and Alexa Fluor594 goat anti-rabbit IgG at 1:250 dilution (A21434 and A11037 respectively, Life Technologies). Eyes from C57BL/6J mice taken 24 h post-infection with 500 colony forming units (c.f.u.) of *Staphylococcus aureus* and from normal C57BL/6J spleen and lymph nodes were used as positive controls for the reported inflammation marker expression studies. The fully human ABCB5 extracellular-loop-associated peptide-targeted IgG1 monoclonal antibody 3B9 was generated by Pfizer Centers for Therapeutic Innovation (CTI) during a collaboration with Boston Children's Hospital using screening of a human scFv phage display library for selection of

binders to biotin-peg2-ABCB5 peptide (RFGAYLIQAGRMTPEG; UniProt accession Q2M3G0), followed by sequence analysis and human IgG1 conversion. The monoclonal antibody 3B9 binds to human ABCB5 peptide (RFGAYLIQAGRMTPEG; UniProt accession Q2M3G0) and mouse Abcb5 peptide (RFGAYLIQAGRMMPPEG; UniProt accession B5X0E4) at concentrations <10 nM, with no significant binding at these concentrations to either scrambled peptide or peptide homologues associated with the related ABCB1 (RFGAYLVAHKLMSPED; UniProt accession P08183), ABCB4 (RFGAYLIVNGHMRFRD; UniProt accession P21439) or ABCB11 (RYGGYLISNEGLHFSY; UniProt accession O95342) proteins. FITC-conjugated ZyMax goat anti-human IgG (H+L) antibody product (817111) was purchased from Invitrogen.

Cell viability assays. Human limbal epithelial cells were seeded on 96-well plates (15,000 per well, $n = 5$ replicates) and treated for 48 h with either blocking concentrations¹² of anti-ABCB5 monoclonal antibody clone 3C2-1D12 (ref. 6) or equivalent concentrations of isotype control monoclonal antibody (clone MOPC31C, Sigma). Cell viability was determined by the CellTiter-Glo Luminescent Cell Viability Assay (Promega) according to the manufacturer's instructions. In additional viability assays employing conditions resembling those used for ABCB5⁺ cell staining for cell isolation, human limbal epithelial cells were incubated on ice with $2 \mu\text{g ml}^{-1}$ of anti-ABCB5 monoclonal antibody clone 3C2-1D12 (ref. 6) or isotype control monoclonal antibody (clone MOPC31C, Sigma) for 30 min. After washing with PBS, cells were seeded on 96-well plate (15,000 per well, $n = 5$ replicates) and grown at 37°C for 48 h, followed by determination of cell viability as described earlier. Statistical differences between samples were determined using the unpaired *t*-test. A two-sided *P* value of $P < 0.05$ was considered significant.

Corneal fragility assay. Experimental animals were anaesthetized with intraperitoneal injections of 70 mg kg^{-1} sodium pentobarbital. Under a stereomicroscope, a partial epithelial defect was created in both eyes by brushing with a wet Microsponge (Alcon) as described by Kao *et al.*²⁵. The animals were euthanized in a CO_2 chamber, and the corneas were removed and embedded in paraffin for histology. H&E-stained *Abcb5* knockout and wild-type corneas were analysed for the presence or absence of epithelial defects and results were compared using the Fisher's exact test. A two-sided *P* value of $P < 0.05$ was considered significant.

Corneal epithelial debridement. After anaesthesia with intraperitoneal injection of ketamine (120 mg kg^{-1} body weight; Hospira) and xylazine (10 mg kg^{-1} body weight; Burns Veterinary Supply), followed by topical application of one drop of 0.5% proparacaine eye drops (Akorn) into each eye, a 2 mm diameter epithelial wound was created by demarcating an area of the central cornea with a 2 mm trephine and removing the epithelium within the circle with a small scalpel, leaving the basement membrane intact. In each animal, the procedure was performed on the right eye. Ak-Spore Ophthalmic Ointment (bacitracin zinc, neomycin sulphate and polymyxin B sulphate; Akorn) was applied to both eyes immediately after wounding and then twice per day for the next 48 h to prevent corneal infection and dryness. Analgesia was provided by subcutaneous injections of Buprenex (Reckitt Benckiser Pharmaceuticals) every 12 h for 48 h post-operatively at the dose of 1 mg kg^{-1} . Wound healing was monitored as described previously²⁶.

Corneal tight junction integrity. Hutcheon *et al.*²⁷ described a functional assay of corneal epithelial cell tight junction integrity using LC-biotin, which does not penetrate through the epithelium in the presence of intact tight junctions, whereas defective tight junctions allow penetration through the epithelium and into the corneal stroma. Wild-type and *Abcb5* knockout mice were assessed for corneal epithelial tight junctions using the LC-biotin staining method performed as described²⁸. In brief, LC-biotin staining solution was prepared by dissolving 1 mg ml^{-1} EZ-Link-Sulfo-NHS-LC-Biotin (Pierce) in Hank's balanced salt solution (HBSS, Lonza) plus 2 mM MgCl_2 and 1 mM CaCl_2 . This solution was applied to the cornea of wild-type and knockout mice for 15 min at the time of euthanasia. Eyes were rinsed with PBS (Lonza), enucleated and placed in Tissue-Tek OCT (Sakura Finetek) for frozen sectioning. Sections of wild-type and knockout corneas were stained with FITC-streptavidin to detect the presence of LC-biotin.

Transplantation experiments. Murine donor limbal epithelial cells were transplanted onto the eyes of syngeneic C57BL/6J recipient mice with an induced limbal stem cell deficiency. Human donor limbal epithelial cells were transplanted onto the eyes of immunodeficient NOD.Cg-Prkdc^{scid} Il2rg^{tm1Wjl}/SzJ (NSG) mice with an induced limbal stem cell deficiency. Four types of donor transplants were performed: (1) ABCB5⁺ limbal epithelial cells; (2) ABCB5⁻ limbal epithelial cells; (3) unsegregated limbal epithelial cells; and (4) grafts containing no cells (fibrin gel carrier only). Twenty-four hours before transplantation, murine and human donor cells were seeded onto a fibrin carrier, which was prepared by dissolving fibrinogen and thrombin stock solutions (TISSUCOL-Kit Immuno, Baxter) in 1.1% NaCl and 1 mM CaCl_2 to a final concentration of 10 mg ml^{-1} fibrinogen and 3 IU ml^{-1} thrombin as described²⁹. On the day of transplantation, total LSCD was induced in anaesthetized recipient mice by removing the corneal and limbal epithelium with an Algerbrush II corneal rust ring remover with a 0.5 mm burr

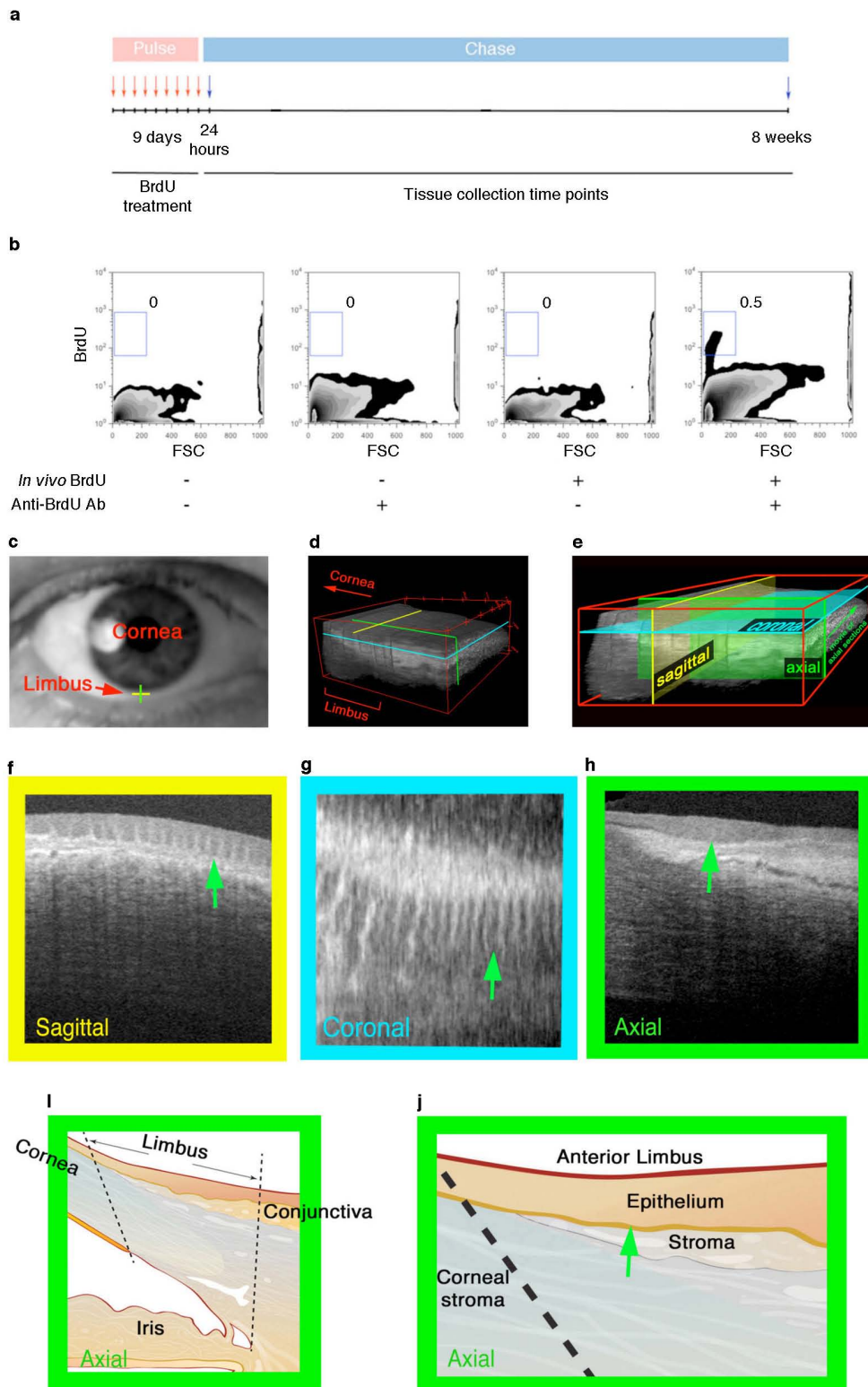
(AMBLER Surgical)³⁰. Following induction of LSCD, recipient mice received fibrin gel carrier-based transplants that were secured via four subconjunctival sutures. Eyelids were sutured with 8-0 nylon sutures to keep the eyes closed. Ak-Spore Ophthalmic Ointment (bacitracin zinc, neomycin sulphate and polymyxin B sulphate; Akorn) was applied on both eyes immediately after wounding and then twice per day for the next 48 h to prevent corneal infection and dryness. Analgesia was provided by subcutaneous injections of $5\text{--}10 \text{ mg kg}^{-1}$ Metacam (Boehringer Ingelheim Pharmaceuticals), given preoperatively and by subcutaneous injections of $0.05\text{--}0.1 \text{ mg kg}^{-1}$ of Buprenex (Reckitt Benckiser Pharmaceuticals) every 12 h for 24 h post-operatively. In addition, after surgical recovery mice were also treated with anti-inflammatory Inflanefran Forte eye drops (Allergan) for the first 5 days, and then with 1% Avastin (Bevacizumab, Genentech) eye drops daily for 5 days. Slit lamp examination was performed weekly until euthanasia. Eyes were enucleated postmortem and fixed in 10% buffered formalin for methacrylate embedding (Technovit, Heraeus Kulzer) or snap-frozen in Tissue-Tek OCT (Sakura Finetek).

Confocal and multiphoton microscopic analysis of corneas. For whole corneal imaging, animals were euthanized and their eyes carefully enucleated. The eyes were then mounted using cyanoacrylate glue with the cornea facing up. Whole corneal imaging was performed using a custom-built video-rate laser-scanning confocal and multiphoton microscope with a custom femtosecond laser supply based on systems described by Veilleux *et al.*³¹ and Wang *et al.*³². In brief, a $1,550 \text{ nm}$ turnkey fibre-based laser (Calmar Laser) with a 5 MHz pulse repetition rate and 360 fs pulse width was frequency doubled using a bismuth borate (BiBO) crystal with an AR1550/775nm coating (Newlight Photonics) to generate $\sim 7.5 \text{ mW}$ at the focal plane. Corneal layers were imaged using confocal reflectance with a quarter wave plate and an avalanche photodiode module (Hamamatsu) to collect back-scattered signal. To further probe the structural elements of the cornea, second harmonic generation of collagen fibrils was collected using a PMT and a $390/40 \text{ nm}$ band pass filter (Semrock). The apex of each cornea was imaged using a $\times 60$ 1.0 NA objective (Olympus) at 15 frames per second with $1 \mu\text{m}$ steps from the surface of the cornea through the basal epithelial layers.

Statistical analysis. We established a provisional number of mice per cohort for the pilot phase of a particular experiment based on initial feasibility considerations with regard to availability of cell graft material, and, if the data were suggestive of a significant difference, we calculated the numbers of mice needed for a repeat experiment to establish statistical significance, or not. There was no exclusion of samples or animals from the analysis. The investigators were blinded when assessing experimental outcomes. In gene knockout analyses, animals were allocated to experimental groups based on their *Abcb5* knockout or *Abcb5* wild-type status. For transplantation experiments all recipient animals were randomly assigned to their respective experimental groups. Two-sided tests were used in the statistical analyses. Appropriate statistical tests were used for all data sets depicted in the figures, with data meeting the assumptions of the tests. Variations within each group of data were estimated and similar between statistically compared groups. * $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$.

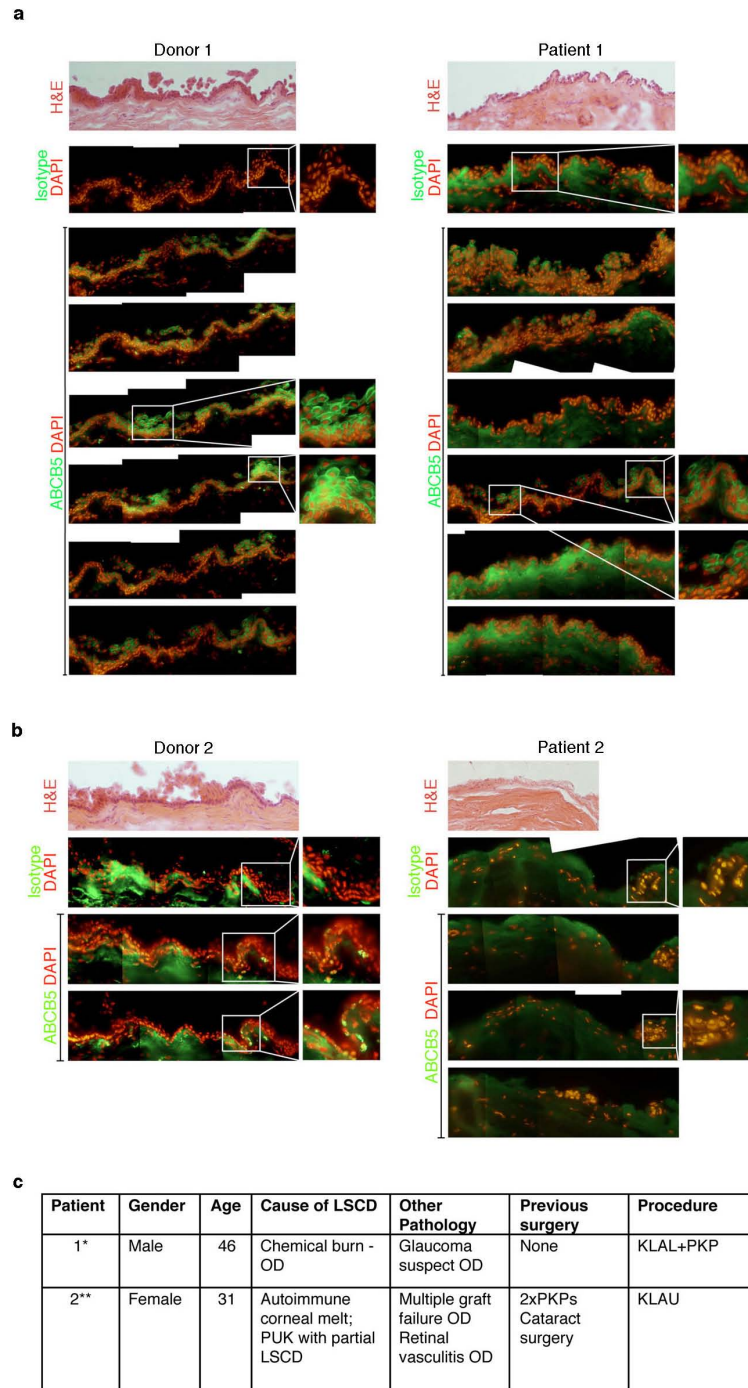
- Liu, P., Jenkins, N. A. & Copeland, N. G. A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Res.* **13**, 476–484 (2003).
- Rodríguez, C. I. *et al.* High-efficiency deleter mice show that FLPe is an alternative to Cre-loxP. *Nature Genet.* **25**, 139–140 (2000).
- Hutcheson, D. A. & Kardon, G. Genetic manipulations reveal dynamic cell and gene functions: Cre-ating a new view of myogenesis. *Cell Cycle* **8**, 3675–3678 (2009).
- Lakso, M. *et al.* Efficient *in vivo* manipulation of mouse genomic sequences at the zygote stage. *Proc. Natl Acad. Sci. USA* **93**, 5860–5865 (1996).
- Frank, N. Y. *et al.* VEGFR-1 expressed by malignant melanoma-initiating cells is required for tumor growth. *Cancer Res.* **71**, 1474–1485 (2011).
- Pellegrini, G. *et al.* Location and clonal analysis of stem cells and their differentiated progeny in the human ocular surface. *J. Cell Biol.* **145**, 769–782 (1999).
- Meyer-Blazejewski, E. A. *et al.* Preservation of the limbal stem cell phenotype by appropriate culture techniques. *Invest. Ophthalmol. Vis. Sci.* **51**, 765–774 (2010).
- Kruloval, M. *et al.* A rapid separation of two distinct populations of mouse corneal epithelial cells with limbal stem cell characteristics by centrifugation on percoll gradient. *Invest. Ophthalmol. Vis. Sci.* **49**, 3903–3908 (2008).
- Lathrop, K. L., Gupta, D., Kagemann, L., Schuman, J. S. & Sundarraj, N. Optical coherence tomography as a rapid, accurate, noncontact method of visualizing the palisades of Vogt. *Invest. Ophthalmol. Vis. Sci.* **53**, 1381–1387 (2012).
- Kao, W. W. *et al.* Keratin 12-deficient mice have fragile corneal epithelia. *Invest. Ophthalmol. Vis. Sci.* **37**, 2572–2584 (1996).
- Pal-Ghosh, S., Tadvalkar, G., Jurjus, R. A., Zieske, J. D. & Stepp, M. A. BALB/c and C57BL/6 mouse strains vary in their ability to heal corneal epithelial debridement wounds. *Exp. Eye Res.* **87**, 478–486 (2008).
- Hutcheon, A., Sippel, K. C. & Zieske, J. D. Examination of the restoration of epithelial barrier function following superficial keratectomy. *Exp. Eye Res.* **84**, 32–38 (2007).
- Klocke, J. *et al.* Spontaneous bacterial keratitis in CD36 knockout mice. *Invest. Ophthalmol. Vis. Sci.* **52**, 256–263 (2011).

29. Pellegrini, G. *et al.* The control of epidermal stem cells (holoclones) in the treatment of massive full-thickness burns with autologous keratinocytes cultured on fibrin. *Transplantation* **68**, 868–879 (1999).
30. Meyer-Blazejewska, E. A. *et al.* From hair to cornea: toward the therapeutic use of hair follicle-derived stem cells in the treatment of limbal stem cell deficiency. *Stem Cells* **29**, 57–66 (2011).
31. Veilleux, I., Spencer, J. A., Biss, D. P., Côté, D. & Lin, C. P. *In vivo* cell tracking with video rate multimodality laser scanning microscopy. *IEEE J. Sel. Top. Quant. Electron.* **14**, 10–18 (2008).
32. Wang, K. *et al.* Three-color femtosecond source for simultaneous excitation of three fluorescent proteins in two-photon fluorescence microscopy. *Biomed. Opt. Express* **3**, 1972–1977 (2012).



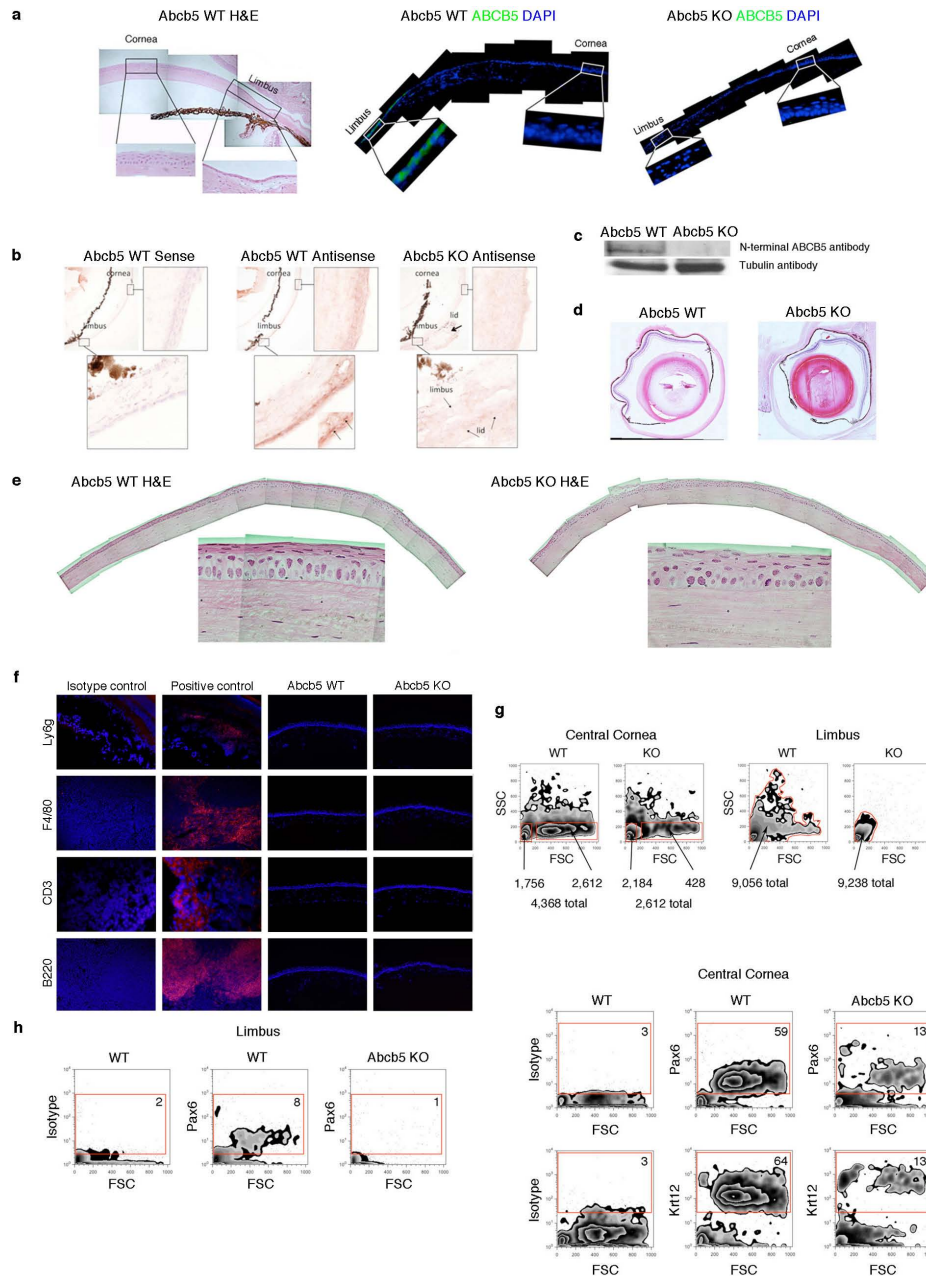
Extended Data Figure 1 | BrdU label-retaining cells and optical coherence tomography identification of the palisades of Vogt. **a**, Schematic summary of the experimental design for BrdU pulse-chase experiments. **b**, Representative flow cytometric analyses depicting specific staining of BrdU label-retaining cells in limbal epithelial cells of wild-type mice that did not receive BrdU (left two panels) or wild-type mice that received BrdU followed by an 8-week chase (right two panels). Limbal epithelial cells were recovered and stained with either anti-BrdU antibody or with an isotype control antibody. The percentages of BrdU-positive cells within the gate are indicated on each plot. **c–e**, Schematic illustration of the optical coherence tomography (OCT) imaging algorithm used for the human limbus. **f–h**, Cross-sectional images of human cornea

depicting a sagittal view (**f**), a coronal C-mode image reconstructed to reveal the palisades (green arrow) and the rete pegs (**g**), and an axial view of the corneolimbal junction showing the conjunctival stroma beneath the limbal epithelium (green arrow identifies the basal epithelial layer) (**h**). **i, j**, Schematic representation of the limbus (**i**) and the anterior limbus (**j**), illustrating the orientation of OCT images used to identify and dissect palisade-rich regions within the limbus of corneal rims (approximately 1 cm² tissue blocks). These smaller sections were then stained with anti-ABCB5 monoclonal antibody and analysed by confocal microscopy, as is shown in Fig. 1e of the main manuscript. Supplementary Video 3 consists of sequential confocal images depicting the location of ABCB5⁺ cells within palisades.



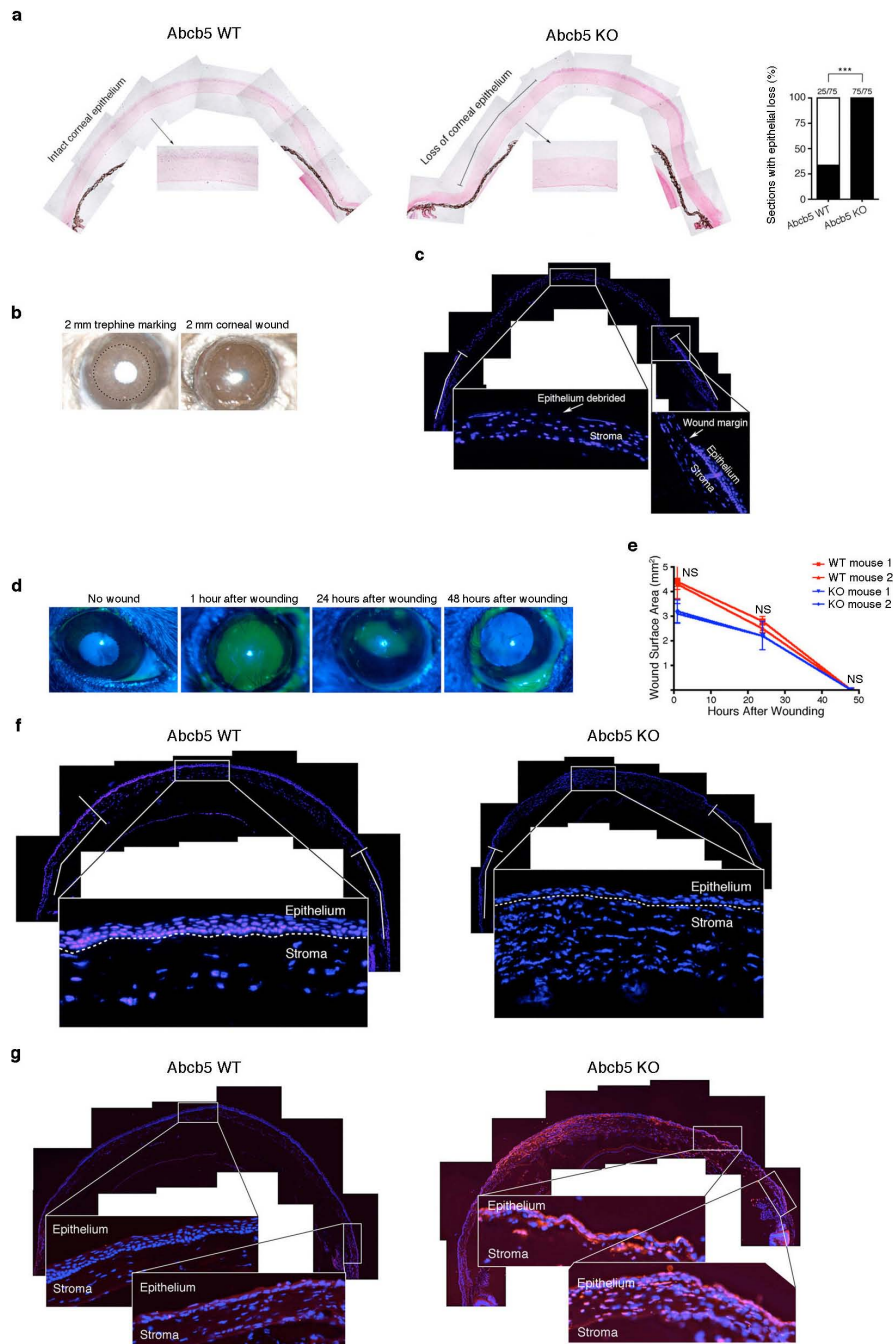
Extended Data Figure 2 | Limbal biopsies from normal donors or patients with LSCD. **a–c**, Limbal biopsies were obtained from normal donors or patients with LSCD. **a**, Typical findings are shown for a patient (patient 1) with a chemical burn before receiving a penetrating keratoplasty plus kerato-limbal allograft from a cadaveric donor eye (donor 1). Serial cross-sections of the biopsies were stained with either H&E, isotype control monoclonal antibody, or ABCB5 monoclonal antibody. ABCB5 staining in the limbal epithelium of donor 1 reveals nests of ABCB5-positive cells, whereas ABCB5-positivity is reduced in the limbal epithelium of patient 1. Photographs of immunofluorescent staining are montages of sequential photos at $\times 20$ magnification. In these studies, equal-sized biopsies were recovered from a portion of the patient and donor limbus, frozen, and sectioned to produce eight sequential sections. All epithelial cells were counted in each section. A total of 2,031 and 2,051 epithelial cells were counted in patient 1 and donor 1, respectively. **b**, Limbal biopsies were obtained from a patient (patient 2) with an autoimmune corneal melt, peripheral ulcerative keratitis, and partial limbal stem cell deficiency before receiving a kerato-limbal autograft from the patient's

normal contralateral eye (donor 2). Serial sections of the biopsies were stained with either H&E, isotype control monoclonal antibody, or ABCB5 monoclonal antibody. ABCB5 positivity was present in the basal layer of the limbal epithelium of donor 2, while a dramatically reduced epithelial layer and no ABCB5 staining was observed in the limbus of patient 2. Photographs of immunofluorescent staining are montages of sequential photos at $\times 20$ magnification. In these studies, equal-sized biopsies were recovered from a portion of the patient and donor limbus, frozen, and sectioned to produce eight sequential sections. All epithelial cells were counted in each section. A total of 563 and 2,662 epithelial cells were counted in patient 2 and donor 2, respectively. Patient 2 had a reduced number of epithelial cells due to the extensive damage from chronic autoimmunity. **c**, LSCD patient information. *Donor 1: cadaveric donor; **Donor 2: autologous transplant from contralateral eye. KLAL, kerato-limbal allograft (limbal tissue was harvested from donor eye); KLAU, kerato-limbal autograft (part of limbal tissue was resected from uninjured contralateral eye); OD, right eye; PKP, penetrating keratoplasty; PUK, peripheral ulcerative keratitis.



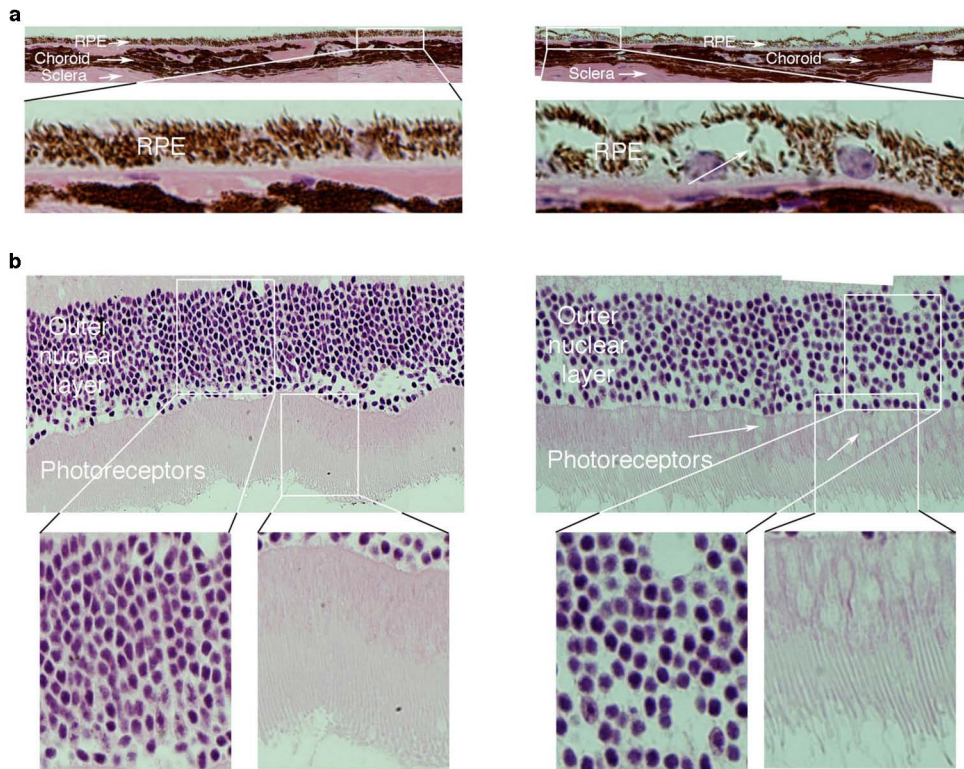
Extended Data Figure 3 | Phenotypic evaluation of *Abcb5* knockout versus wild-type mice. **a**, H&E staining of the normal murine eye depicts location of the limbus and central cornea (left panel). Representative immunofluorescence staining of the wild-type (WT) murine eye illustrates the presence of an *Abcb5*⁺ cell population (green) in the limbus but not the central cornea (middle panel). *Abcb5* immunofluorescence staining of the knockout (KO) mouse demonstrates loss of *Abcb5* expression in the limbus (right panel). **b**, *In situ* hybridization with murine antisense or sense *Abcb5* mRNA probes spanning 144 bp of murine *Abcb5* cDNA encoding exon 10 of the murine *Abcb5* gene (GenBank accession number JQ655148) reveals loss of *Abcb5* mRNA expression in *Abcb5* knockout mice. **c**, Western blot analyses of murine protein lysates with rabbit polyclonal antibody directed against the *Abcb5* N terminus (Abgent) reveal loss of a ~80 kDa protein band of the predicted size in *Abcb5* knockout mice. **d**, H&E cross-section of a wild-type eye (left) and an *Abcb5* knockout eye (right) depicts a normal shape of the *Abcb5* knockout eye and the presence of all major structures including cornea, conjunctiva, iris, lens and retina. **e**, H&E staining of methacrylate-embedded *Abcb5* wild-type (left) and *Abcb5* knockout (right) age-matched adult corneas. **f**, Inflammatory cell marker immunostaining and respective isotype control immunostaining in positive control tissues (left columns) and inflammatory cell marker immunostaining of *ABCB5* wild-type and *ABCB5* knockout corneas (right

columns). Ly-6G, neutrophil marker; F4/80, macrophage marker; CD3, T-cell marker; B220, B-cell marker. Positive control tissues: *Staphylococcus aureus*-infected murine cornea (Ly-6G); murine spleen (F4/80 and B220); murine lymph node (CD3). **g**, Representative flow cytometry analyses of either the central corneal (left) or the limbal (right) epithelium of wild-type and *Abcb5* knockout mice. Forward scatter (FSC) and side scatter (SSC) indicate cellular size and granularity, respectively. The central corneal epithelium of *Abcb5* knockout mice showed reduced numbers of epithelial cells compared to wild-type epithelium (left panels), caused by a reduction in larger cells (right gates), but not smaller cells (left gates). There was no reduction in the numbers of limbal epithelial cells (right panels). Representative results of samples pooled from four eyes are shown ($n = 3$ experiments). **h**, Representative flow cytometry analyses of epithelial cells harvested from either the limbus or the central cornea of wild-type and *Abcb5* knockout mice. Recovered cells were stained with isotype control antibody, anti-Pax6 antibody, or anti-Krt12 antibody. There was a reduced frequency of Pax6⁺ cells in the limbus of *Abcb5* knockout mice and a reduced frequency of Pax6⁺ and Krt12⁺ epithelial cells in the central cornea of *Abcb5* knockout mice. Red gates identify Pax6⁺ or Krt12⁺ cells compared to isotype control staining. Representative plots of $n = 3$ experiments for each marker are shown. Magnification in **a**, **b**, **e**, **f**: $\times 20$; **d**: $\times 4$.



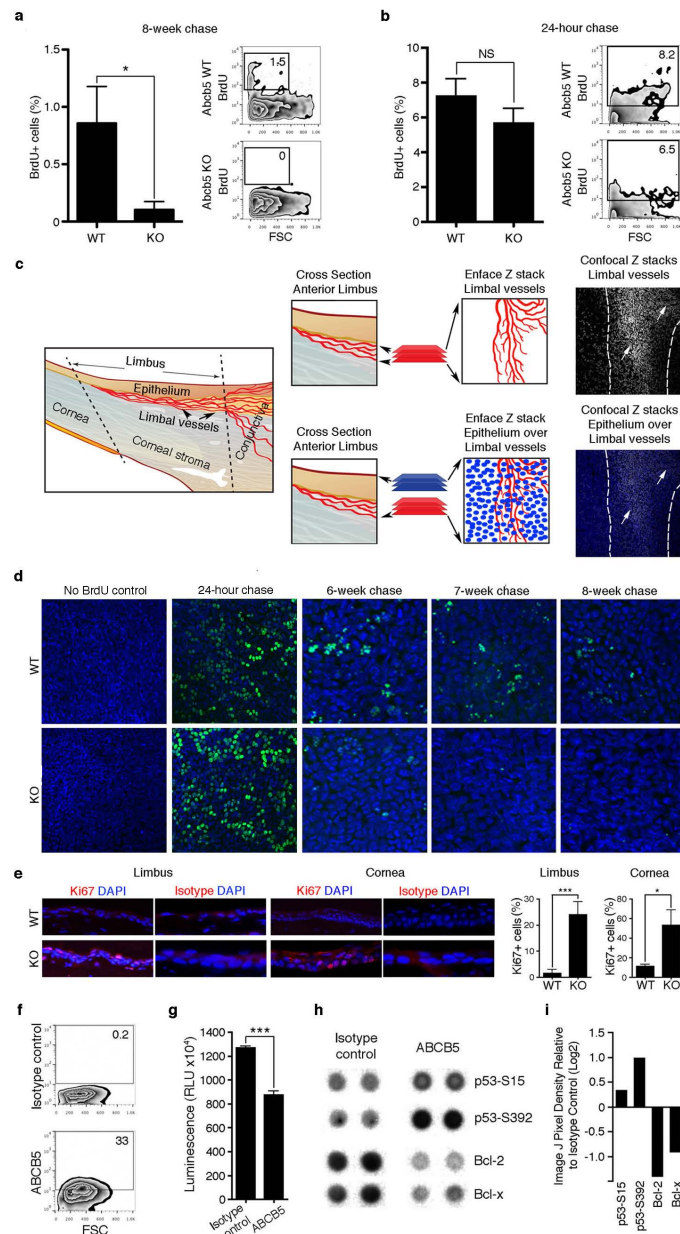
Extended Data Figure 4 | Functional evaluation of *Abcb5* knockout versus wild-type mice. **a**, Increased corneal fragility in *Abcb5* knockout mice. H&E-stained sections of wild-type (WT) and *Abcb5* knockout (KO) corneas collected immediately after brushing with a wet Microsponge were examined for the presence or absence of epithelial defects. Only 33% of wild-type animal-derived cornea sections exhibited small epithelial defects (<25% of epithelium), whereas 100% of *Abcb5* knockout cornea sections exhibited significant epithelial injury ($n = 3$ mice per group, 25 sections per mouse, Fisher's exact test: $P < 0.001$). **b–e**, Wound healing following corneal epithelial debridement of wild-type and *Abcb5* knockout mice. **b**, The area to be debrided was marked with a 2 mm trephine and the epithelium was removed with a small scalpel. **c**, DAPI-stained cross-section of the cornea immediately following central epithelial debridement depicting the wound margins and exposed central corneal stroma. Image is a montage of sequential photos at $\times 10$ magnification. **d**, Corneal epithelial wound closure was monitored at 1, 24, and 48 h post-debridement via fluorescein staining. **e**, Wound closure rates were not significantly different between wild-type and *Abcb5* knockout mice (summary of $n = 2$ replicate experiments, $n = 4$ mice per group, unpaired t -test, P = not significant). **f**, Reduced re-epithelialization of wounded corneas

in *Abcb5* knockout mice. Representative DAPI-stained composite corneal cross sections of wild-type (left) and *Abcb5* knockout (right) mice 48 h after a corneal epithelial debridement wound, demonstrating a reduced number of epithelial cells in *Abcb5* knockout mice. The white dashed line demarcates the epithelium from stroma; the white box indicates area shown at $\times 20$ magnification (montage pictures are at $\times 10$ magnification); white lines demarcate the area in which epithelial cells were counted. Epithelial cells were counted within the standardized area in at least three consecutive composite cross sections in three replicate mice per group in two separate experiments (aggregate data shown in Fig. 2i). **g**, Increased apoptosis in wounded corneas of *Abcb5* knockout mice. Representative TUNEL-stained composite corneal cross-sections of wild-type (left) and *Abcb5* knockout (right) mice 48 h after a corneal epithelial debridement wound, demonstrating increased numbers of apoptotic cells in *Abcb5* knockout mice. Areas defined by the white box are shown at $\times 20$ magnification (montage pictures at $\times 10$ magnification). The number of TUNEL-positive epithelial cells was counted, and the data from two replicate experiments in $n = 2$ mice are summarized in Fig. 2k. Error bars indicate s.e.m. NS, not significant.



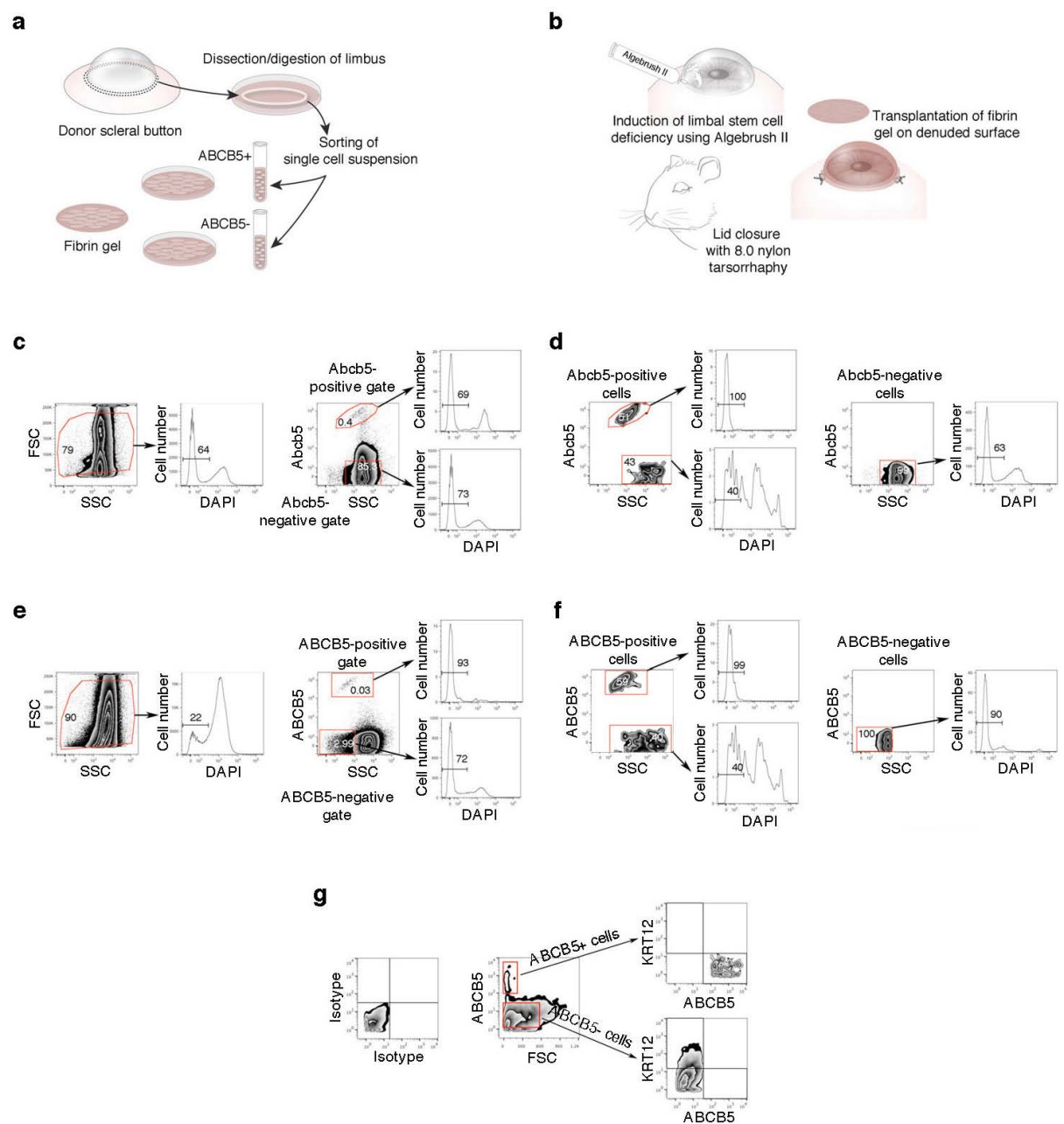
Extended Data Figure 5 | Phenotypic characterization of wild-type versus *Abcb5* knockout retina. Analysis of H&E-stained sections from 9-month-old wild-type mice (left) or *Abcb5* knockout mice (right) revealed changes in the retina and retinal pigment epithelial cells (RPEs). **a**, Compared to wild-type mice (left), RPEs in *Abcb5* knockout mice (right) were enlarged and distended, possibly due to the presence of vacuoles. **b**, Compared to wild-type mice (left),

areas of abnormal RPE in *Abcb5* knockout eyes (right) coincided with changes in the overlaying photoreceptors and the outer nuclear layer. There was a thinning and attenuation of photoreceptor outer segments along with a disruption of inner segments, which was associated with a loss of cells in the outer nuclear layer. Magnification: $\times 20$.



Extended Data Figure 6 | ABCB5 regulates LSC quiescence and functions as an anti-apoptotic molecule. **a**, Flow cytometric analysis showing depletion of BrdU label-retaining cells in *Abcb5* knockout (KO) versus *Abcb5* wild-type (WT) limbal epithelial cells after an 8-week chase. Analysis was performed in $n = 6$ *Abcb5* wild-type mice and in $n = 12$ *Abcb5* knockout mice. The experiment was performed three times. Data were analysed using the unpaired t -test. Data are shown as mean \pm s.e.m., $P < 0.05$. **b**, Equivalent BrdU labelling in *Abcb5* wild-type and *Abcb5* knockout mice after a 24 h chase is shown (mean \pm s.e.m.). The experiment was performed using $n = 4$ mice per group and was performed twice. Data were analysed using the unpaired t -test, $P =$ not significant. **c**, Schematic illustrating identification of the limbal epithelium within corneal whole mounts via identification of posteriorly localized limbal vessels within the underlying stroma. Far right, confocal Z-stack images displaying limbal vessels alone (top, white arrows) and limbal vessels (white arrows) with overlying limbal epithelium (bottom) (see also Supplementary Video 2). **d**, Sequential immunofluorescence histological examination of limbal epithelium in corneal whole mounts (localized as illustrated in c and Supplementary Video 2), showing equivalent BrdU incorporation after a 24 h chase in *Abcb5* knockout and wild-type mice (column 2), but progression to selective loss of BrdU label-retaining cells in *Abcb5* knockout mice at the 8-week chase time point (far right column). Far left column, negative BrdU

immunostaining result (negative control) of BrdU-untreated wild-type and *Abcb5* knockout mouse limbal epithelium. **e**, Immune fluorescence analysis of Ki67 expression in *Abcb5* wild-type and *Abcb5* knockout mouse limbus and cornea. Bar graphs on the right illustrate the percentage of Ki67+ cells in *Abcb5* wild-type versus *Abcb5* knockout mice in the limbus and cornea. The percentages of Ki67+ cells in *Abcb5* wild-type versus *Abcb5* knockout mice in the limbus and cornea were determined using $n = 4$ mice per group. The experiment was performed twice. Within a standardized area, all corneal epithelial cells were counted in at least three consecutive cross-sections. Data were analysed using the unpaired t -test. Data are shown as mean \pm s.e.m. **f**, Flow cytometric analysis of ABCB5 expression by p63 α -rich human limbal epithelial cells. **g**, Cell viability measured in relative luciferase units (RLU) following 48 h of ABCB5 monoclonal antibody or isotype control monoclonal antibody treatment ($n = 5$ experimental replicas, mean \pm s.e.m.). Data were analysed using the unpaired t -test, $P < 0.001$. **h**, **i**, Differential expression of apoptosis pathway-associated proteins detected by Proteome Profiler Apoptosis Array (ARY009, R&D Systems) following ABCB5 monoclonal antibody or isotype control monoclonal antibody treatment, analysed using ImageJ software. Error bars indicate s.e.m. * $P < 0.05$, *** $P < 0.001$. NS, not significant. Magnification in **c–e**: $\times 20$.

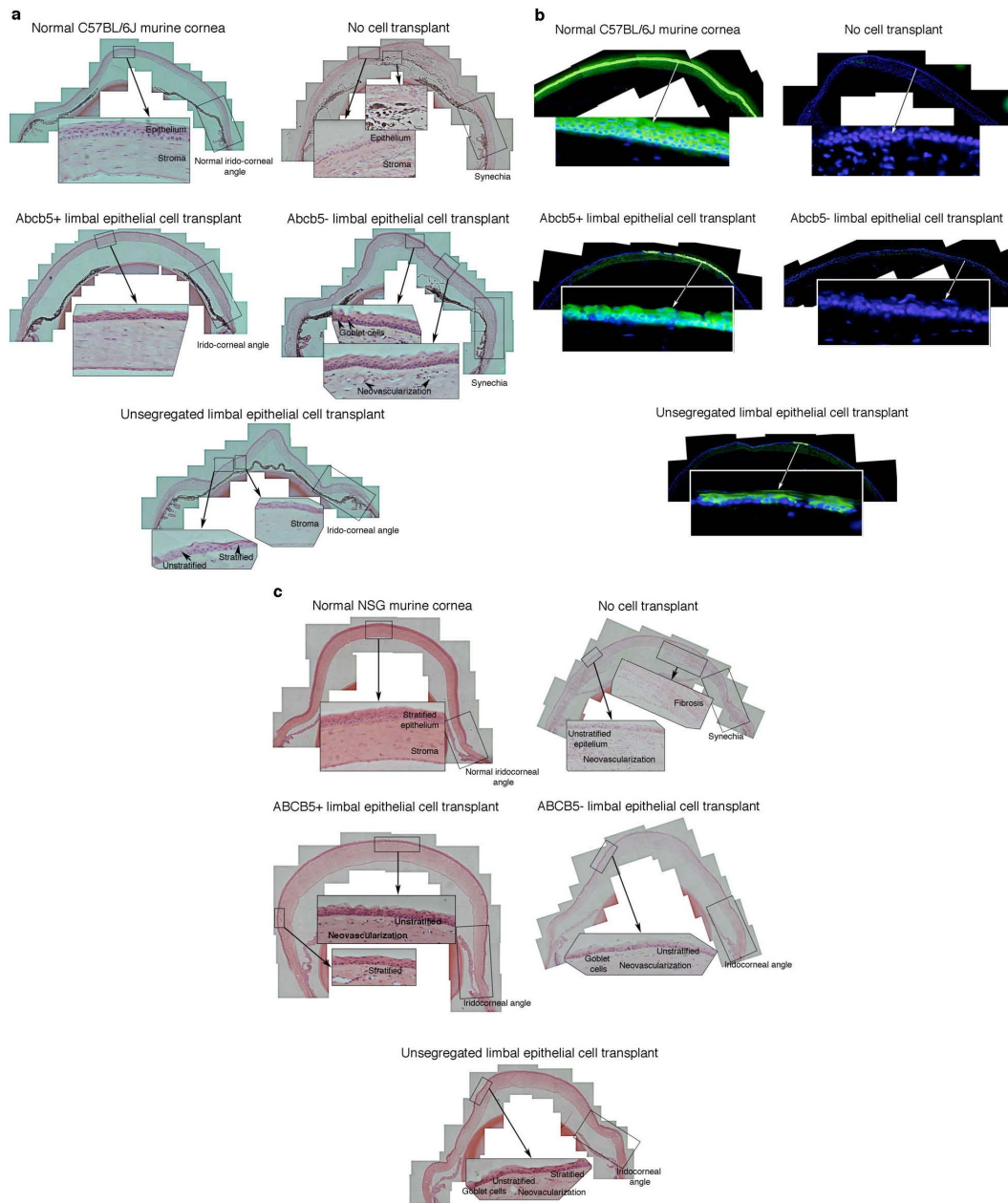


h

Donor	Unsegregated grafts 500 cells/graft/mouse		ABCB5(+) enriched grafts 500 cells/graft/mouse		ABCB5(-) enriched grafts 500 cells/graft/mouse	
Mouse limbus	ABCB5		ABCB5		ABCB5	
	Positive	Negative	Positive	Negative	Positive	Negative
Viable cells/graft	1	319	255	86	0	312
Human limbus	ABCB5		ABCB5		ABCB5	
	Positive	Negative	Positive	Negative	Positive	Negative
Viable cells/graft	1	109	292	80	0	450

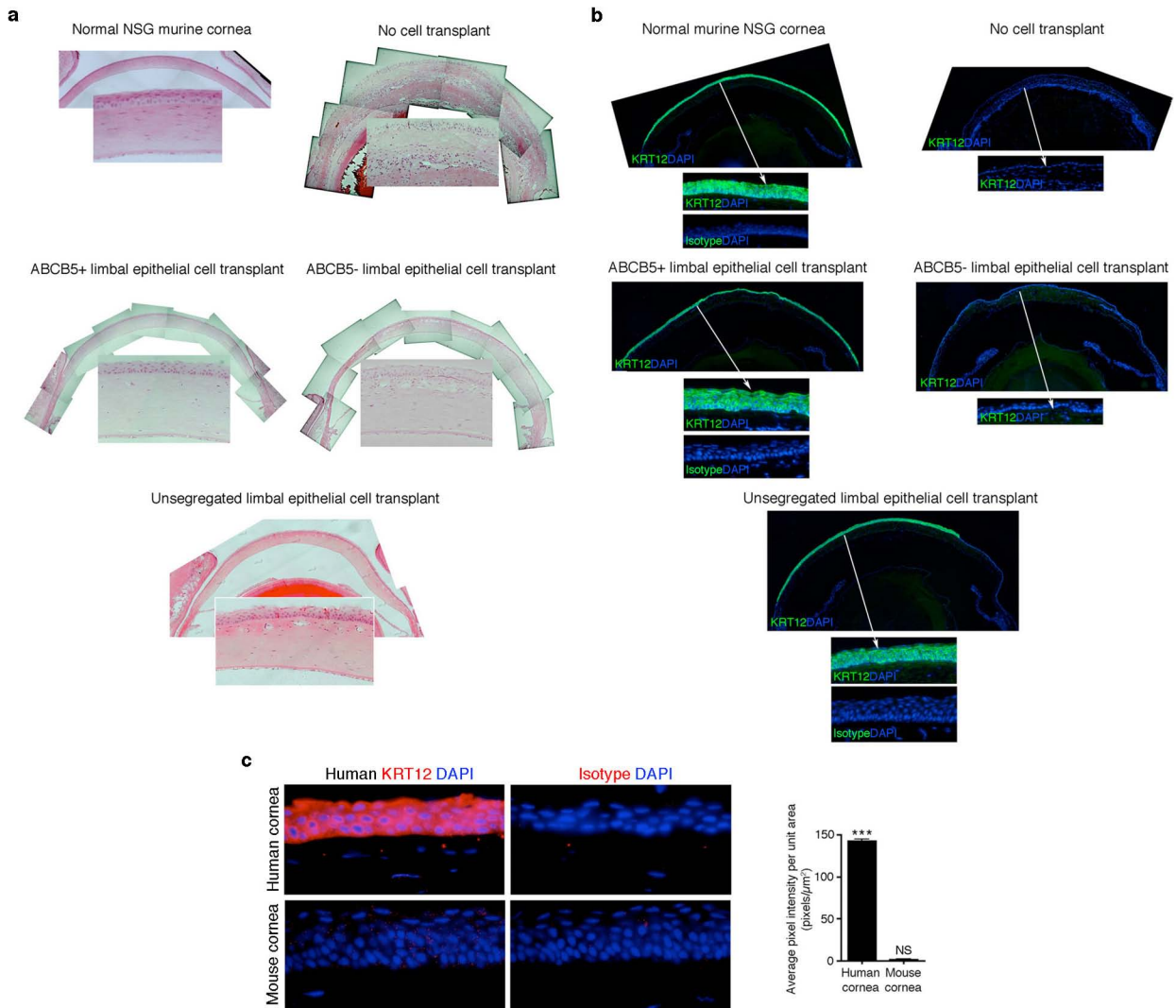
Extended Data Figure 7 | Limbal stem cell transplantation protocol and cell sorting for purification of ABCB5⁺ and ABCB5⁻ limbal epithelial cells. **a**, Recovery and separation of ABCB5⁺ and ABCB5⁻ limbal epithelial cells from donor corneas followed by preparation of fibrin gels containing donor cells. **b**, Induction of limbal stem cell deficiency in recipient mice and transplantation of donor grafts. **c**, **e**, Representative flow cytometry analyses showing sorting gates and viability of murine (c) and human donor limbal

epithelial cells (e). **d**, **f**, Post-sort analysis depicting the purity and viability of ABCB5⁺-enriched and ABCB5⁻-enriched subpopulations of limbal epithelial cells isolated from mice (d) and human donors (f). Viability is shown as the percentage of cells excluding DAPI. **g**, KRT12 expression in ABCB5⁺ and ABCB5⁻ limbal cell populations. **h**, Number and viability of donor cells used for transplantation.



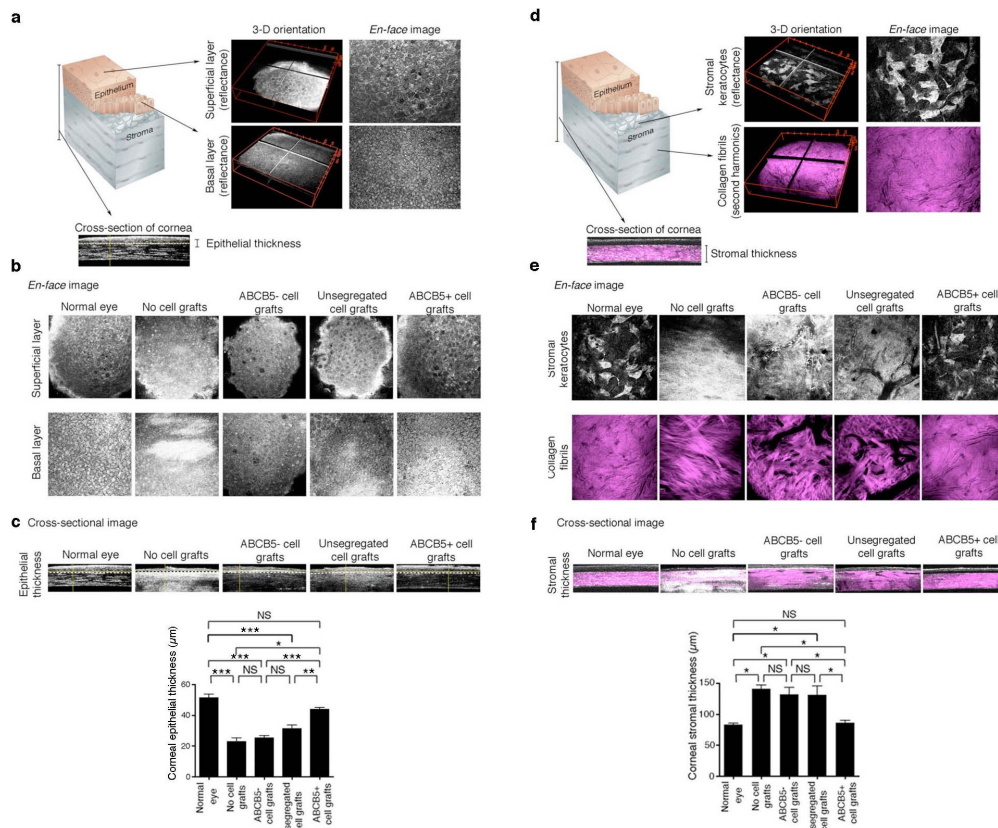
Extended Data Figure 8 | Restoration of LSCD by donor murine *Abcb5*⁺ or human *ABCB5*⁺ cell transplants. Representative H&E composite corneal cross-sections of recipient C57BL/6J mice 5 weeks after receiving an induced LSC deficiency followed by engraftment of donor fibrin gel transplants containing the following syngeneic murine limbal epithelial cell subpopulations: (1) no cells (negative control), (2) *Abcb5*⁺ cells, (3) *Abcb5*⁻ cells, or (4) unsegregated cells. A normal untreated cornea (no LSCD) served as a positive control. The positive control displays the typical stratified corneal epithelium and iridocorneal angle. Mice receiving transplants with no cells displayed the typical conjunctivalization that occurs following a LSCD, that is, unstratified conjunctival epithelium covers the cornea with extensive inflammation, neovascularization, and stromal oedema. Synechia (where the iris adheres to the cornea) is typical of intense anterior segment inflammation. In contrast, mice that received transplants of *Abcb5*⁺ cells, but not *Abcb5*⁻ cells, displayed a restored stratified corneal epithelium with no evidence of inflammation, neovascularization, stromal oedema, or synechia. Mice that received transplants of unsegregated limbal epithelial cells displayed areas of stromal oedema with unstratified epithelium, while other parts of the cornea contained normal stratified epithelial cells. **b**, Restoration of Krt12 expression by donor murine *Abcb5*⁺ cell transplants. Representative immunofluorescent Krt12 staining (green) of recipient C57BL/6J mice 5 weeks after an LSCD induction followed by transplantation of donor fibrin gel grafts containing

grafts as in **a**. Normal untreated murine cornea (no LSCD), shown here as a positive control, displays high intensity of Krt12 staining. Mice that received grafts containing no cells displayed no Krt12 expression. In contrast, mice transplanted with *Abcb5*⁺ cells, exhibited significantly enhanced Krt12 expression in comparison to the mice transplanted with unsegregated limbal epithelial cells. No Krt12 expression was detected in the mice transplanted with *Abcb5*⁻ cells. The white box depicts the area shown at $\times 40$ magnification. Montage images are shown at $\times 10$ magnification. **c**, Restoration of LSCD by donor human *ABCB5*⁺ limbal cell transplants. Representative H&E composite corneal cross-sections of recipient immunodeficient NSG mice 5 weeks after LSCD induction followed by transplantation of donor fibrin gel grafts containing the following human limbal epithelial cell subpopulations: (1) no cells (negative control), (2) *ABCB5*⁺ cells, (3) *ABCB5*⁻ cells, and (4) unsegregated cells. The positive control (normal untreated NSG cornea (no LSCD)) displays the typical stratified corneal epithelium and iridocorneal angle. Mice that received transplants with no cells displayed evidence of conjunctivalization that occurs following a LSC deficiency; that is, unstratified conjunctival epithelium covers the cornea with extensive neovascularization and synechia (anterior segment inflammation is muted in NSG mice due to their immunodeficiency). In contrast, mice that received transplants containing *ABCB5*⁺ cells displayed areas of restored stratified epithelium, whereas recipients receiving *ABCB5*⁻ cell grafts did not.



Extended Data Figure 9 | Long-term corneal restoration by donor human ABCB5⁺ cell transplants 13 months after transplantation. **a**, Representative H&E composite corneal cross-sections of recipient immunodeficient NSG mice 13 months after LSCD induction followed by transplantation of donor fibrin gel grafts containing: (1) no cells (negative control), (2) ABCB5⁺ cells, (3) ABCB5⁻ cells, and (4) unsegregated cells. A normal untreated NSG cornea (no LSCD) served as a positive control. The positive control displays the typical stratified corneal epithelium. Mice that received transplants with no cells displayed evidence of conjunctivalization that occurs following a LSC deficiency; that is, unstratified conjunctival epithelium covers the cornea with extensive neovascularization and synechia (anterior segment inflammation is muted in NSG mice due to their immunodeficiency). In contrast, mice that received transplants containing ABCB5⁺ cells displayed restored stratified epithelium, whereas recipients receiving ABCB5⁻ cell grafts did not. **b**, Representative immunofluorescent KRT12 staining (green) of corneas derived from NSG mice 13 months after LSCD induction followed by transplantation of donor fibrin gel grafts containing the following human limbal epithelial cell subpopulations: (1) no cells (negative control), (2) ABCB5⁺ cells, (3) ABCB5⁻ cells, or (4) unsegregated cells. Normal untreated murine cornea (no LSCD), shown here as a positive control,

displayed a high intensity of KRT12 staining. As expected, recipients of grafts containing no cells displayed no KRT12 expression. Mice transplanted with ABCB5⁺ cells exhibited significant KRT12 expression, also enhanced compared to mice transplanted with unsegregated limbal epithelial cells. No KRT12 expression was detected in mice transplanted with ABCB5⁻ cells. The white arrow depicts the area shown at $\times 40$ magnification. Montage images are shown at $\times 10$ magnification. **c**, Representative immunofluorescent KRT12 staining (red) of human and mouse cornea confirms specific antibody reactivity with human KRT12 (top left), and no cross-reactivity with murine Krt12 (bottom left). Isotype control antibody staining is shown for the respective tissues in the right panels. Nuclei are stained with DAPI (blue). Bar graph (bottom) demonstrates aggregate antibody staining data of either human cornea (pixel intensity 142.3 ± 2.4 pixels μm^{-2} , mean \pm s.e.m.) or mouse cornea (pixel intensity 1.3 ± 0.7 pixels μm^{-2} , mean \pm s.e.m.). Aggregate human KRT12 antibody staining data of either human or mouse cornea was derived from the analyses on $n = 2$ corneas per group. Within a standardized area in (b), all corneal epithelial cells were counted in at least $n = 3$ consecutive cross-sections. Data were analysed using the unpaired *t*-test. Error bars show s.e.m. ****P* < 0.001. NS, not significant.



Extended Data Figure 10 | Corneal epithelial restoration in 13-month-old human transplants. **a–c**, Corneal epithelial restoration in 13-month-old human transplants examined by reflectance confocal microscopy. **a**, Schematic illustration of the cornea. **b**, **c**, Confocal microscopic reflectance was used to image $400 \times 400 \mu\text{m}$ areas in the central cornea in the normal eye, as well as control and treatment groups. Representative *en face* images of normal corneal epithelial layers depicting superficial squamous and basal cuboidal epithelial cells, and the corresponding cross-sectional image depicting the epithelial layer thickness, are shown in the extreme left panels of **b** and **c**, respectively.

Additional panels in **b** and **c** from left to right show, show representative *en face* (**b**) and cross-sectional (**c**) images of recipient NSG mice 13 months after LSCD induction followed by transplantation of donor fibrin gel grafts containing the following human limbal epithelial cell subpopulations: (1) no cells (negative control), (2) ABCB5⁻ cells, (3) unsegregated cells, or (4) ABCB5⁺ cells. Normal untreated cornea (no LSCD), shown here as a positive control, displayed a typical stratified epithelium of normal thickness. Mice that received grafts containing no cells displayed no stratified epithelium and a significantly reduced epithelial layer. Mice transplanted with ABCB5⁻ cells displayed a thin unstratified epithelium that was not significantly different from the negative control group. Mice transplanted with unsegregated limbal cells displayed a mixture of stratified and unstratified epithelium that was significantly thinner compared to normal corneas. In contrast, only mice transplanted with ABCB5⁺ cells displayed a normal stratified epithelium with superficial squamous and basal cuboidal epithelial cells, and a thickness not only significantly greater than in alternative treatment or untreated control groups, but comparable to normal healthy cornea (no significant difference), as determined by measurements of cross-sectional image data. Epithelial layer thickness measurements were performed in all groups using ImageJ software and cross-sectional reflectance confocal microscopy imaging (4 mice per group, 10 measurements per cornea) through the central region of the cornea. The measurements were performed on mice from two independent experiments. Data were analysed using the one-way ANOVA and Bonferroni multiple comparisons tests. Aggregate results are illustrated in **c**, bottom panel bar graph (mean thickness in micrometres (μm) \pm s.e.m.).

d–f, Corneal stromal architecture of 13-month-old human transplants examined by reflectance confocal and second harmonic generation microscopy. **d**, Schematic illustration of the cornea. **e**, **f**, Reflectance confocal and second harmonic

generation microscopy was used to image $400 \times 400 \mu\text{m}$ areas in the central stroma of the normal eye, as well as of control and treatment groups. Representative *en face* images of normal cornea (**e**, extreme left panels) show normal stromal keratocytes as determined by confocal reflectance (top) and stromal architecture as determined by second harmonic generation of collagen fibrils (magenta images, bottom). The corresponding cross-sectional image depicting the stroma layer thickness is shown in the extreme left of panel **f**. Additional panels in **e** and **f** from left to right show representative *en face* and cross-sectional images of recipient NSG mice 13 months after LSCD induction followed by transplantation of donor fibrin gel grafts containing the following human limbal epithelial cell subpopulations: (1) no cells (negative control), (2) ABCB5⁻ cells, (3) unsegregated cells, or (4) ABCB5⁺ cells. Normal untreated cornea (no LSCD), shown here as a positive control, displayed typical stromal keratocytes and a normal collagen fibril pattern, with normal stromal thickness determined in cross-sectional images. Mice that received grafts containing no cells displayed a high level of reflectance due to inflammation (also compare H&E staining in Extended Data Fig. 9a) and stromal oedema as shown by increased stromal thickness. In addition, an abnormal collagen fibril pattern was observed, possibly due to deposition of new collagen by infiltrating inflammatory cells. Mice transplanted with ABCB5⁻ cells displayed a high level of reflectance, an abnormal collagen fibril pattern, and stromal oedema that was not significantly different from the negative control group. Mice transplanted with unsegregated limbal cells also displayed increased reflectance, an abnormal collagen fibril pattern, and stromal oedema. In contrast, only mice transplanted with ABCB5⁺ cells displayed a normal pattern of stromal keratocytes and collagen fibrils, and a stromal thickness comparable to normal healthy cornea (no significant difference) and indicative of absent oedema, as determined by measurements of cross-sectional image data. Stromal thickness measurements were performed in all groups using ImageJ software and cross-sectional second harmonic microscopic images of collagen fibrils (4 mice per group, 5 measurements per stroma) through the central region of the stroma. The measurements were performed on mice from two independent experiments. Data were analysed using the one-way ANOVA and Bonferroni multiple comparisons tests. Aggregate results are illustrated in **f**, bottom panel bar graph (mean thickness in μm \pm s.e.m.). Error bars show s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. NS, not significant. Magnification in **a–f** $\times 60$.

WNT7A and PAX6 define corneal epithelium homeostasis and pathogenesis

Hong Ouyang^{1,2}, Yuanchao Xue³, Ying Lin^{1,2}, Xiaohui Zhang^{2†}, Lei Xi^{1,2}, Sherrina Patel², Huimin Cai^{4,5}, Jing Luo², Meixia Zhang⁴, Ming Zhang⁴, Yang Yang^{2†}, Gen Li⁴, Hairi Li³, Wei Jiang⁴, Emily Yeh², Jonathan Lin², Michelle Pei², Jin Zhu², Guiqun Cao⁴, Liangfang Zhang^{2,6}, Benjamin Yu^{7,8}, Shaochen Chen^{2,6}, Xiang-Dong Fu^{2,3,8}, Yizhi Liu¹ & Kang Zhang^{1,2,4,8,9}

The surface of the cornea consists of a unique type of non-keratinized epithelial cells arranged in an orderly fashion, and this is essential for vision by maintaining transparency for light transmission. Cornea epithelial cells (CECs) undergo continuous renewal from limbal stem or progenitor cells (LSCs)^{1,2}, and deficiency in LSCs or corneal epithelium—which turns cornea into a non-transparent, keratinized skin-like epithelium—causes corneal surface disease that leads to blindness in millions of people worldwide³. How LSCs are maintained and differentiated into corneal epithelium in healthy individuals and which key molecular events are defective in patients have been largely unknown. Here we report establishment of an *in vitro* feeder-cell-free LSC expansion and three-dimensional corneal differentiation protocol in which we found that the transcription factors p63 (tumour protein 63) and PAX6 (paired box protein PAX6) act together to specify LSCs, and WNT7A controls corneal epithelium differentiation through PAX6. Loss of WNT7A or PAX6 induces LSCs into skin-like epithelium, a critical defect tightly linked to common human corneal diseases. Notably, transduction of PAX6 in skin epithelial stem cells is sufficient to convert them to LSC-like cells, and upon transplantation onto eyes in a rabbit corneal injury model, these reprogrammed cells are able to replenish CECs and repair damaged corneal surface. These findings suggest a central role of the WNT7A–PAX6 axis in corneal epithelial cell fate determination, and point to a new strategy for treating corneal surface diseases.

Corneal and skin epithelium share many similarities, including a typical morphology of stratified epithelium and maintenance of their stem cells by p63 in the keratin 5/keratin 14⁺ (K5/K14)-expressing basal cell layer in limbus and epidermis^{4–8} (Fig. 1a, b and Extended Data Fig. 1a, b). However, there are marked differences between them. Skin epithelial stem cells (SESCs) move upwards from a deep to suprabasal layers vertically during differentiation^{9,10}, where K5 and K14 are replaced by skin-specific K1 and K10 (ref. 11 and Extended Data Fig. 1c, d). In contrast, LSCs (defined by K19 at the limbus¹², see Fig. 1a and Extended Data Fig. 1e) migrate centripetally for several millimetres to the central cornea during which it undergoes differentiation and K5/K14 are replaced by corneal-specific K3 and K12 (refs 13, 14, Fig. 1c and Extended Data Fig. 1f).

A clear, transparent cornea maintained by CECs is essential for vision. Pathological conversion of CECs into skin-like epithelial cells, as indicated by morphological changes and switches in keratin expression (for example, replacement of cornea-specific K3 and K12 by skin-specific K1 and K10 along with K5⁺ cells at the basal layer; see Fig. 1d), leads to the loss of transparency in the cornea and causes millions of people around the world to suffer from partial or complete blindness³, but the underlying mechanism has remained largely unknown.

To elucidate potential disease mechanisms, we successfully developed a feeder-free cell culture protocol to expand LSCs from human donors, enabling us to generate a homogeneous cell population to delineate key factors involved in controlling LSC cell fate determination and CEC differentiation. Proliferating LSCs were characterized by positive p63 and K19 with a high percentage of mitotic marker Ki67 (Fig. 2a and Extended Data Fig. 1g). We next established a three-dimensional LSC differentiation protocol to establish a three-dimensional CEC sphere structure from a single LSC within 14 to 18 days, as evidenced by strong expression of the CEC-specific markers K3 and K12 (Fig. 2b). The three-dimensional differentiation sphere was further characterized by key differences in gene expression between LSCs and CECs; the latter showed increased expression of K3 (31.2-fold higher) and K12 (24.7-fold higher)

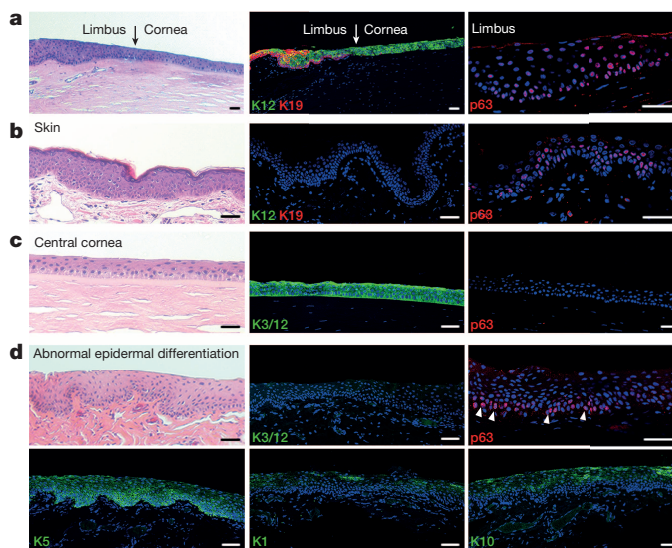


Figure 1 | Normal and pathological changes of corneal epithelium, and its comparison to skin. **a**, Normal cornea–limbus junction (arrows). Limbus identified by K19 and p63 (also see Extended Data Fig. 1e), and cornea by K12. **b**, Normal skin epidermis identified by p63 and K5/K14 (see Extended Data Fig. 1a, b) in the basal layer and absence of K3 and K12 (K3/12). **c**, Normal central cornea labelled by K3/12 and absence of p63 and K1 and K10 (see also Extended Data Fig. 1c, d, f). **d**, Cornea with abnormal epidermal differentiation showing absence of K3/12 (top middle panel) and presence of skin epithelium makers p63 (top right panel) and K5, K1 and K10 (bottom panels). Haematoxylin and eosin (H&E) staining was used for the left panels, with the exception of the bottom left panel. Scale bars, 100 μ m.

¹State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou 510060, China. ²Department of Ophthalmology, and Biomaterial and Tissue Engineering Center of Institute of Engineering in Medicine, University of California San Diego, La Jolla, California 92093, USA. ³Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, California 92093, USA. ⁴Molecular Medicine Research Center, State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, Sichuan 610041, China. ⁵Guangzhou KangRui Biological Pharmaceutical Technology Company Ltd., Guangzhou 510005, China. ⁶Department of Nanoengineering, University of California San Diego, La Jolla, California 92093, USA. ⁷Department of Medicine, University of California San Diego, La Jolla, California 92093, USA. ⁸Institute for Genomic Medicine, University of California San Diego, La Jolla, California 92093, USA. ⁹Veterans Administration Healthcare System, San Diego, California 92093, USA. [†]Present addresses: Beijing Institute of Ophthalmology, Beijing Tongren Eye Center, Beijing 100730, China (X.Z.); Department of Ophthalmology, Shengjing Hospital of China Medical University, Shenyang 110004, China (Y.Y.).

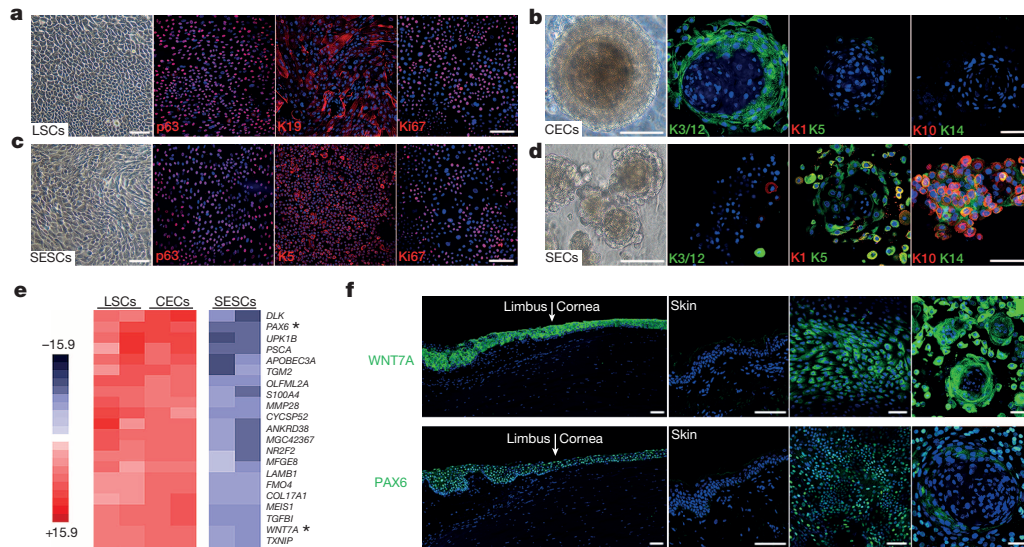


Figure 2 | Exclusive expression of WNT7A and PAX6 at limbus and cornea. **a–d**, Immunofluorescence staining of cultured LSCs and SECs, and three-dimensional differentiated CECs and SECs. Left panels, phase contrast photographs; staining of p63, K19 and Ki67 in LSCs (**a**), p63, K5 and Ki67 in SECs (**c**), K3/12, K1, K5, K10 and K14 in CECs (**b**) and SECs (**d**) in three-dimensional culture spheres. **e**, Heatmap depicting differential gene expression

and concomitant decreased expression of *K19* (6.2-fold lower, all $P < 0.01$; see Extended Data Fig. 1h). We took a similar strategy to expand SECs and observed strong expression of typical SEC markers p63 and K5 in cultured SECs (Fig. 2c). As expected, we detected increased expression of epidermal differentiation markers *K1* (16.6-fold higher) and *K10* (225.8-fold higher) in three-dimensional differentiated skin epithelial cells (SECs) compared to SECs (Fig. 2d, Extended Data Fig. 1i, j).

To identify additional genes uniquely expressed in LSCs, CECs and SECs, we performed genome-wide gene expression analysis (Fig. 2e and Extended Data Fig. 2a, b). Among genes that were differentially expressed, we focused on signalling molecules and transcription factors because of their central roles in cell fate determination and differentiation. We identified that *WNT7A* and *PAX6* were highly expressed in LSCs and CECs when compared to SECs (*PAX6*, 8.8-fold higher in LSCs and 12.3-fold higher in CECs, $P < 0.001$; *WNT7A*, 4.5-fold higher in LSCs, 6.0-fold higher in CECs, $P < 0.001$) (Fig. 2e and Extended Data Fig. 2c). We observed that *WNT7A* expression precisely mirrored the expression pattern of *PAX6* in *in vitro* LSC and CEC cultures, and in *in vivo* epithelial layers of cornea and limbus from infant to adult, while both of these genes were undetectable in skin epidermis (Fig. 2f and Extended Data Fig. 2d).

To determine the clinical relevance of *WNT7A* and *PAX6* expression in LSCs and CECs, we examined several types of human corneal diseases, corneal epithelium squamous metaplasia, inflammatory keratopathy, trauma and alkaline burn. We observed the localized expression of p63 and K5 at the basal layer (Fig. 3a and Extended Data Fig. 3), and the expression of K10 in the suprabasal layer (Fig. 1d and Extended Data Fig. 3). We also found that *WNT7A* and *PAX6* expression, and K3 and K12 expression were conspicuously absent in areas of metaplasia, while they were positive in surrounding corneal epithelium (Fig. 3a and Extended Data Fig. 3). These results suggest cornea epithelial cells were switched to skin-like epithelial cells in patient tissues with these disease conditions.

Wnt molecules are secreted signalling proteins that have a critical role in controlling cell fate decisions and tissue specification¹⁵. *PAX6* is also a well-known control gene for eye development and disease¹⁶. However, it has remained unclear whether the loss of *PAX6* is the cause or the consequence of abnormal skin epidermal differentiation in ocular surface diseases.

comparing among LSCs, CECs and SECs. Asterisks indicate *WNT7A* and *PAX6*. **f**, Immunofluorescence staining of *WNT7A* and *PAX6* at limbus, cornea and skin (left and middle left panels). Expression of *WNT7A* and *PAX6* in cultured LSCs (middle right panels) and three-dimensional CEC spheres (right panels). Scale bars, 100 μ m.

To demonstrate that *WNT7A* and *PAX6* are necessary for LSC and CEC cell fate determination and differentiation, we used lentiviral short hairpin RNAs (shRNAs) to knock them down specifically in LSCs. Although LSCs with knockdown of either *WNT7A* or *PAX6* did not change proliferation and morphological properties (Extended Data Fig. 4a), these treatments significantly diminished the expression of corneal K3

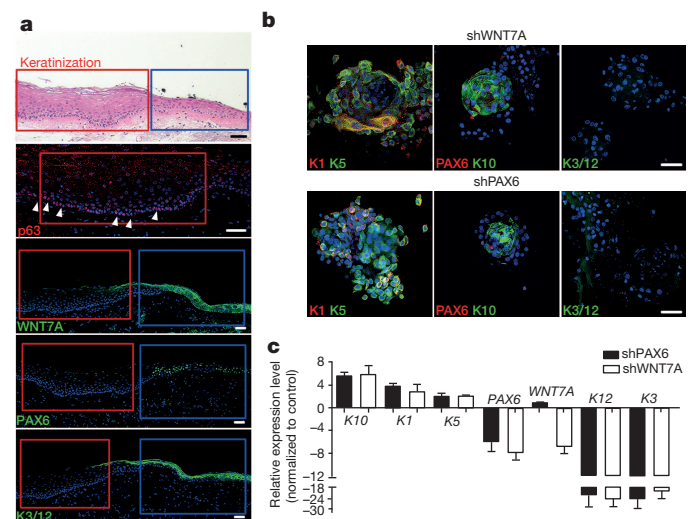


Figure 3 | WNT7A and PAX6 are essential for maintenance of cornea cell fate. **a**, Human corneal epithelium squamous metaplasia. In the top panel, the red box indicates the area of metaplasia and the blue box indicates the area of relatively normal cornea. H&E stain (top panel) shows typical skin epidermal morphology with p63⁺ at basal layer (second panel, arrowheads indicate p63 staining). Loss of *WNT7A* (middle panel) and *PAX6* (fourth panel) was accompanied by absence of corneal K3/12 (bottom panel). Serial sections of the areas marked by red and blue boxes in the top panel are represented in the lower panels. **b**, Immunofluorescence of three-dimensional differentiated cells with *WNT7A* or *PAX6* knockdown; left panels, K1 and K5; middle panels, *PAX6* and K10; right panels, K3/12; **c**, Quantitative PCR analysis of gene expression changes of cornea or skin epithelium markers in three-dimensional differentiated cells with *WNT7A* or *PAX6* knockdown (all $n = 3$, $P < 0.05$). Data are shown as means \pm s.d. Scale bars, 100 μ m.

and K12 under the three-dimensional differentiation conditions (*WNT7A* knockdown: 24.7-fold lower in K3, 22.6-fold lower in K12; *PAX6* knockdown: 20.8-fold lower in K3, 21.4-fold lower in K12; all $P < 0.05$), and concurrently, the expression of skin-specific K1 and K10 became more prominent (*WNT7A* knockdown: 3.9-fold higher in K1 and 5.7-fold higher in K10; *PAX6* knockdown: 3.1-fold higher K1 and 6.1-fold higher in K10; all $P < 0.05$), indicative of more skin-like differentiation (Fig. 3b, c). Moreover, knockdown of *WNT7A* reduced *PAX6* expression in LSCs (1.8-fold lower, $P < 0.001$); this repressive effect was even stronger in differentiated CECs (8.0-fold lower, $P < 0.01$). In contrast, there was no significant difference in *WNT7A* expression when *PAX6* was knocked down in either LSCs or CECs (Fig. 3c and Extended Data Fig. 4b, c). These results suggest that *WNT7A* acts upstream of *PAX6* during CEC differentiation.

To study further the role of the Wnt signalling pathway in corneal fate determination and differentiation, we investigated the functional requirement of Frizzled receptors, which have been shown to interact and transduce *WNT7A* signalling based on co-immunoprecipitation¹⁷. We found that *WNT7A* interacted strongly with Frizzled 5 (*FZD5*) in LSCs (Extended Data Fig. 4d, e), and as predicted, knockdown of *FZD5* in LSCs also led to reduced *PAX6* expression (1.7-fold lower in LSCs and 3.0-fold lower in differentiated CECs ($P < 0.001$) (Extended Data Fig. 4f). Together, these data demonstrated that loss of *WNT7A* or *PAX6* led to a switch of corneal epithelial cells to skin-like epidermal

cells and that *WNT7A* and *FZD5* acted as the upstream regulators of *PAX6* expression in corneal differentiation.

Given the central role of *PAX6* in eye development¹⁶, we next tested the possibility that engineered expression of *PAX6* might be able to convert SENCs into LSC-like cells (Extended Data Fig. 5a). Indeed, we found that the expression of either *PAX6a* or *PAX6b* in SENCs was sufficient to convert them into LSC-like cells, as evidenced by the induced K19 expression on the surface, coincident with the expression of both p63 and *PAX6* in the nucleus (Fig. 4a). When placed in three-dimensional culture, *PAX6*-transduced SENCs showed dramatic increase in corneal K3 and K12 expression (9.4-fold higher and 72.7-fold higher, all $P < 0.05$) with concomitant decrease in skin K1 and K10 expression (20.8-fold lower and 20.0-fold lower, all $P < 0.01$) (Fig. 4b, c and Extended Data Fig. 5b, c). To obtain global evidence for successful cell fate conversion, we performed gene expression profiling by RNA sequencing (RNA-seq)¹⁸ on CECs, SENCs and LSCs after knocking down *PAX6* and on SENCs transduced with *PAX6* upon three-dimensional differentiation. We generated 3 to 7 million reads from each biological sample that were uniquely

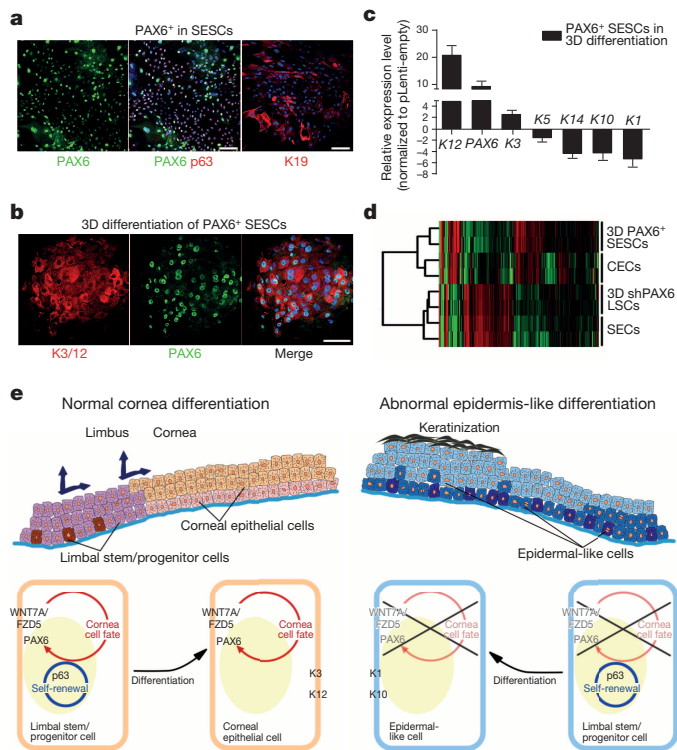


Figure 4 | Conversion of SENCs into corneal epithelial-like cells by *PAX6* transduction. **a**, Double immunofluorescence staining of *PAX6* and p63 in transfected SENCs, K19 was positive in *PAX6*-transduced (*PAX6*⁺) SENCs. **b**, Immunofluorescence staining of K3/12 and *PAX6*⁺ SENCs in three-dimensional (3D) differentiation conditions. **c**, QPCR analysis of gene expression of keratins in *PAX6*⁺ SENCs (all $n = 3$, $P < 0.05$). Data are shown as means ± s.d. **d**, Hierarchical cluster analysis among CECs, differentiated LSCs with *PAX6* knockdown (three-dimensional shPAX6 LSCs), SECs and differentiated SENCs with *PAX6* transduction (three-dimensional *PAX6*⁺ SENCs). **e**, Schematic diagram showing normal LSCs differentiation into CECs (left panel) and proposed mechanism in which loss of *WNT7A*/*PAX6* in LSCs leads to abnormal skin epidermis-like differentiation in corneal surface epithelial cell disease (right panel). Scale bars, 100 μm.

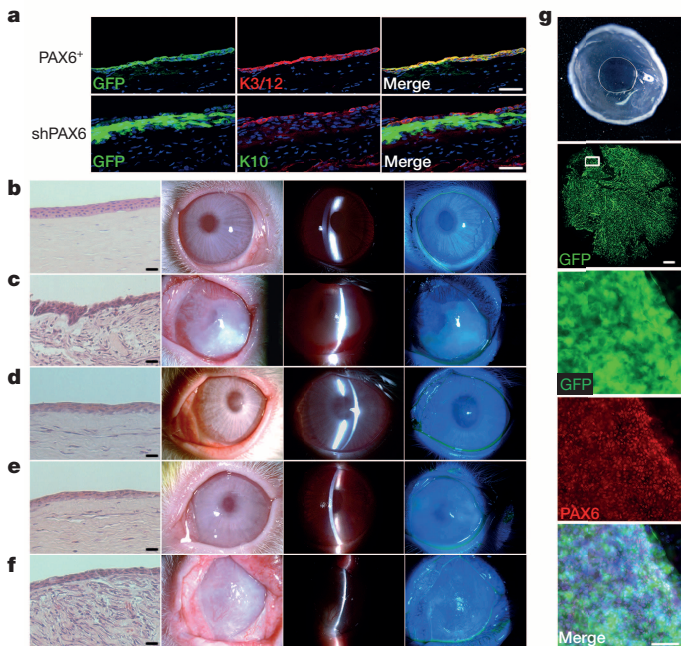


Figure 5 | Cell transplantation and cornea epithelium repair in a rabbit limbal stem cell deficiency model. **a**, Immunofluorescence staining of rabbit corneas 2 months post transplantation. Top panels, cornea transplanted with GFP-labelled *PAX6*⁺ SENCs, showing positive GFP signals and the expression of the corneal epithelium markers K3 and K12 on the corneal surface. Bottom panels, cornea transplanted with GFP-labelled shPAX6 LSCs, showing positive GFP signals and the expression of the skin epidermal epithelium marker K10. Scale bars, 100 μm. **b–f**, Rabbit corneas 2 months post cell transplantation (left panels, H&E stain; middle two panels, white light micrograph and slit-lamp micrograph; right panels, fluorescein dye staining of corneal epithelium surface). Scale bars, 100 μm. **b**, Normal cornea with typical corneal epithelium histology and intact cornea surface without epithelial defects. **c**, Denuded cornea covered with a human amniotic membrane only, showing histology of epithelial metaplasia and opaque cornea with vascularization ($n = 4$). **d, e**, Cornea transplanted with GFP-labelled LSCs (**d**, $n = 3$) and GFP-labelled *PAX6*⁺ SENCs (**e**, $n = 5$), showing corneal epithelium histology, healed and intact cornea surface without epithelial defects. **f**, Cornea transplanted with GFP-labelled, shPAX6-treated LSCs, showing histology of epithelial metaplasia, opaque and vascularized corneal surface with epithelial defects ($n = 4$). **g**, Rabbit cornea 3 months post transplantation with GFP-labelled *PAX6*⁺ SENCs: smooth, transparent cornea (top panel) with positive GFP signals (second panel, scale bar, 1 mm). The framed area in the second panel is enlarged to show the expression of *PAX6* (middle, fourth and bottom panels, scale bar, 100 μm).

mapped to the RefSeq database (Extended Data Fig. 6a). Pairwise comparison demonstrated that the data were very reproducible within the same group of samples (Extended Data Fig. 6b); in contrast, when compared between cells with different fates, the data demonstrate remarkable differences based on the statistical cut-off of false discovery rate (FDR) < 0.001 (Extended Data Fig. 6c). We displayed the entire data sets that recorded the expression of $> 10,000$ genes in various cell types (Fig. 4d), demonstrating that both induced (red) and repressed (green) genes were clearly co-segregated between CECs and PAX6⁺ SESC and between PAX6 shRNA-treated LSCs and SECs. These data therefore provided global evidence for a role of the WNT7A–PAX6 axis in cell fate conversion from SESC to CECs. Together, these data suggest that defects in the WNT7A–PAX6 axis are likely to be responsible for metaplastic conversion of corneal cells to skin epidermal-like cells in corneal diseases in humans (shown in Fig. 4e), although further studies need to be performed to determine the significance of the WNT7 and PAX6 axis in corneal epithelial differentiation.

Finally, we tested the treatment and repair potential of SESC with engineered expression of PAX6 (Extended Data Fig. 7a–c) for corneal epithelial defects in a rabbit LSC deficiency model (Extended Data Fig. 7f), which mimics a common corneal disease condition in humans. We showed that rabbit SESC with PAX6 transduction formed a continuous sheet of epithelial cells with positive staining of corneal-specific K3 and K12 (Fig. 5a) and successfully repaired epithelium defect of the entire corneal surface to restore and maintain normal cornea clarity and transparency for over 3 months (Fig. 5b–g and Extended Data Fig. 8). By following the time course of corneal epithelial surface repair using GFP-labelled PAX6⁺ SESC, we observed that these PAX6-reprogrammed SESC were initially only located at the limbal region and then moved progressively towards the central cornea with corresponding areas of restored cornea clarity (Extended Data Fig. 9a). Importantly, these grafted cells were indeed able to repopulate limbus as evidenced by culture and re-isolation of PAX6⁺ SESC from limbal region (Extended Data Fig. 9b). Notably, these reprogrammed SESC were capable of repairing large corneal epithelium defects after repeated corneal epithelial scraping (Extended Data Fig. 9c). In marked contrast, transplanting rabbit LSCs with PAX6 knockdown (Extended Data Fig. 7a, d, e) onto denuded corneal surface resulted in a K10⁺ skin-like epithelium with opacity and vascularization (Fig. 5f). Together, these data demonstrate that SESC with PAX6 expression are able to trans-differentiate into corneal-like epithelium and repair corneal surface defects.

In summary, this work establishes the feasibility of expanding LSC under feeder-free conditions and its therapeutic potential, and demonstrates key roles of WNT7A and PAX6 in corneal lineage specification. Importantly, SESC or other cell types converted into a corneal fate by PAX6 expression may serve as a potential source for corneal surface repair and regeneration, particularly in patients with total LSC deficiency. This would overcome a major feasibility problem in using a patient's own reprogrammed LSCs for transplantation, thus pointing to a potential therapeutic strategy for treating many common corneal diseases in humans.

METHODS SUMMARY

LSCs and SESC were isolated from rabbits and human donors in feeder-free media and differentiated in the three-dimensional culture conditions. Histology, immunohistochemistry and immunocytochemistry were carried out on paraffin sections as well as on cultured cells. Gene expression microarray, RNA-seq and quantitative PCR (qPCR) were performed using total RNA isolated from LSCs, SESC and CECs.

Lentiviral RNA interference and engineered-expression study of WNT7A, PAX6 and FZD5 were carried out in LSCs and SESC. Cell transplantation of LSCs and SESC was performed on animal models of corneal injury. Detailed information is provided in the supplement.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 August 2013; accepted 12 May 2014.

Published online 2 July 2014.

1. Davanger, M. & Evensen, A. Role of the pericorneal papillary structure in renewal of corneal epithelium. *Nature* **229**, 560–561 (1971).
2. Cotsarelis, G., Cheng, S. Z., Dong, G., Sun, T. T. & Lavker, R. M. Existence of slow-cycling limbal epithelial basal cells that can be preferentially stimulated to proliferate: implications on epithelial stem cells. *Cell* **57**, 201–209 (1989).
3. Li, W. *et al.* Down-regulation of Pax6 is associated with abnormal differentiation of corneal epithelial cells in severe ocular surface diseases. *J. Pathol.* **214**, 114–122 (2008).
4. Pellegrini, G. *et al.* p63 identifies keratinocyte stem cells. *Proc. Natl Acad. Sci. USA* **98**, 3156–3161 (2001).
5. Mills, A. A. *et al.* p63 is a p53 homologue required for limb and epidermal morphogenesis. *Nature* **398**, 708–713 (1999).
6. Yang, A. *et al.* p63 is essential for regenerative proliferation in limb, craniofacial and epithelial development. *Nature* **398**, 714–718 (1999).
7. Koster, M. I., Kim, S., Mills, A. A., DeMayo, F. J. & Roop, D. R. p63 is the molecular switch for initiation of an epithelial stratification program. *Genes Dev.* **18**, 126–131 (2004).
8. Rama, P. *et al.* Limbal stem-cell therapy and long-term corneal regeneration. *N. Engl. J. Med.* **363**, 147–155 (2010).
9. Blanpain, C. & Fuchs, E. Epidermal homeostasis: a balancing act of stem cells in the skin. *Nature Rev. Mol. Cell Biol.* **10**, 207–217 (2009).
10. Arwert, E. N., Hoste, E. & Watt, F. M. Epithelial stem cells, wound healing and cancer. *Nature Rev. Cancer* **12**, 170–180 (2012).
11. Kopan, R. & Fuchs, E. A new look into an old problem: keratins as tools to investigate determination, morphogenesis, and differentiation in skin. *Genes Dev.* **3**, 1–15 (1989).
12. Lauweryns, B., van den Oord, J. J. & Missotten, L. The transitional zone between limbus and peripheral cornea. An immunohistochemical study. *Invest. Ophthalmol. Vis. Sci.* **34**, 1991–1999 (1993).
13. Eichner, R., Bonitz, P. & Sun, T. T. Classification of epidermal keratins according to their immunoreactivity, isoelectric point, and mode of expression. *J. Cell Biol.* **98**, 1388–1396 (1984).
14. Schlötzer-Schrehardt, U. & Kruse, F. E. Identification and characterization of limbal stem cells. *Exp. Eye Res.* **81**, 247–264 (2005).
15. Dorsky, R. I., Moon, R. T. & Raible, D. W. Control of neural crest cell fate by the Wnt signalling pathway. *Nature* **396**, 370–373 (1998).
16. Eiraku, M. *et al.* Self-organizing optic-cup morphogenesis in three-dimensional culture. *Nature* **472**, 51–56 (2011).
17. von Maltzahn, J., Bentzinger, C. F. & Rudnicki, M. A. Wnt7a-Fzd7 signalling directly activates the Akt/mTOR anabolic growth pathway in skeletal muscle. *Nature Cell Biol.* **14**, 186–191 (2012).
18. Yu, F.-X. *et al.* Regulation of the Hippo-YAP pathway by G-protein-coupled receptor signalling. *Cell* **150**, 780–791 (2012).

Acknowledgements This study is supported in part by the 973 program (2013CB967504 and 2014CB964900), Project of Fundamental Research Funds (no.2012KF03), State Key laboratory of Ophthalmology, NIH (GM049369), KACST-UCSD Center of Excellence in Nanomedicine, NIH Director's Transformative R01 Program (R01 EY021374) and CIRM.

Author Contributions H.O., X.-D.F., Yiz.L., and K.Z. designed study, interpreted data and wrote the manuscript. H.O., Y.X., Yin.L., X.Z., L.X., H.C., J.L., Mei.Z., Min.Z., Y.Y., H.L., G.L., E.Y., G.C., J.Z. and B.Y. performed the experiments. Y.L., W.J., J.L. and Yiz.L. obtained human samples. S.C., S.P., M.P. and L.Z. contributed to data analysis and interpretation.

Author Information Microarray and RNA sequence information has been submitted to the Gene Expression Omnibus database under accession number GSE32145 and GSE54322. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.Z. (kangzhang@ucsd.edu), X.-D.F. (xdffu@ucsd.edu) or Y.Z.L. (yizlu62@yahoo.com).

METHODS

Human pathology samples. Corneal epithelium squamous metaplasia and all other tissues were obtained as de-identified surgical specimens, fixed in 5% formalin, embedded into paraffin, sectioned and stained for immunofluorescence studies.

Isolation and culture of limbal stem cells and skin epidermal stem cells. Post-mortem human eyeballs were obtained from eye banks and limbus regions were taken and washed in cold PBS with 100 international units (IU) penicillin and $100 \mu\text{g ml}^{-1}$ streptomycin, and cut into small pieces. Cell clusters were obtained by 0.2% collagenase IV digestion at 37°C for 2 h, single cells were obtained by further digestion with 0.25% trypsin-EDTA at 37°C for 15 min. Primary cells were seeded on plastic plates coated with 2% growth factor reduced Matrigel (354230, BD Biosciences). Limbal stem cells from GFP-labelled rats and rabbits were isolated and cultured using the same method as for human LSCs.

Human epidermis was obtained from donor skin biopsy of eye lids, and hair follicles were removed under microscope. Primary human and rabbit epidermal stem cells were isolated from interfollicular epidermis using the same method as described for human limbal stem cells. Culture medium was as follows: DMEM/F12 and DMEM (1:1) with 1/100 penicillin-streptomycin, 10% fetal bovine serum, 10 ng ml^{-1} EGF, $5 \mu\text{g ml}^{-1}$ insulin, $0.4 \mu\text{g ml}^{-1}$ hydrocortisone, 10^{-10} M cholera toxin and 2×10^{-9} M 3,3',5-triiodo-L-thyronine.

All cells used in the current manuscript are from primary cultured cells made in our laboratories, and mycoplasma contamination tests were routinely carried out and were negative.

In vitro three-dimensional differentiation protocol. Three-dimensional differentiation was performed on a 24-well plate or an 8-well chamber. In brief, dissociated single stem cells were embedded in matrigel at 2×10^4 cells per $50 \mu\text{l}$ gel. Three-dimensional structures were formed after 14–18 days culture in a differentiation medium CnT-30 (limbal stem cell differentiation) or CnT-02 (skin epidermal stem cell differentiation) (Cellntec).

Immunofluorescence and laser confocal microscopy. To detect the localization of proteins in cultured cells, cells were fixed with 4% paraformaldehyde for 20 min, then permeabilized with 0.3% Triton X-100-PBS for 5 min twice and blocked in PBS solution containing 5% bovine serum albumin and 0.3% TritonX-100%, followed by an overnight incubation in primary antibodies at 4°C . After three washes in PBS, cells were incubated with secondary antibody. Cell nuclei were counterstained with DAPI (4',6-diamidino-2-phenylindole). For immunofluorescence of paraffin-embedded tissue sections, de-paraffinization was performed, followed by the same immunofluorescence protocol described above.

The following antibodies were used: mouse anti-p63 monoclonal antibody, rabbit anti-K5 monoclonal antibody, mouse anti-K10 monoclonal antibody, mouse anti-K14 monoclonal antibody with biotin labelled, mouse anti-K19 monoclonal antibody, (MA1-21871, RM210650, MS611P0, MS115B0, MS1902P0, Thermo Fisher Scientific), rabbit anti-PAX6 polyclonal antibody (PRB-278P, Covance), mouse anti-K1 monoclonal antibody (sc-376224, Santa Cruz), Rabbit anti-WNT7A polyclonal antibody, mouse anti-K3/K12 monoclonal antibody, rabbit anti-K12 monoclonal antibody (ab100792, ab68260, ab124975, Abcam), mouse anti-Ki67 monoclonal antibody (550609, BD Biosciences), anti-GFP rabbit monoclonal antibody and anti-GFP mouse monoclonal antibody (G10362, A11120, Invitrogen). The secondary antibodies, AlexaFluor-488- or 568-conjugated anti-mouse or rabbit immunoglobulin-G (IgG) (Invitrogen) were used at a dilution of 1:500. Images were obtained using an Olympus FV1000 confocal microscope.

Quantitative PCR. RNA was isolated using an RNeasy kit (Qiagen) and subjected to on-column DNase digestion. Complementary DNA synthesis was performed using a superscript III reverse transcriptase kit according to the manufacturer's instructions (Invitrogen). qPCR was performed by 40-cycle amplification using gene-specific primers (Extended Data Table 1; top) and a Power SYBR Green PCR Master Mix on a 7500 Real Time PCR System (Applied Biosystems). Measurements were performed in triplicates and normalized to endogenous GAPDH levels. Relative fold change in expression was calculated using the $\Delta\Delta\text{CT}$ method (cycle threshold (CT) values < 30). Data are shown as mean \pm s.d. based on three replicates.

Genome-wide gene expression microarray and data analysis. Total RNA was isolated from LSCs, SENCs and differentiated CECs from three-dimensional differentiation assay. Gene expression microarray analysis was performed using an Illumina human genome microarray system, with each sample in biological replicate ($n = 2$ per group; Human HT-12 v4 Expression BeadChip; Illumina, San Diego, California). Raw data were deposited into the GEO database under accession number GSE32145. Expression-level data were generated by the Illumina BeadStudio version 3.4.0 and normalized using quartile normalization. Probes whose expression level exceeded a threshold value of 64 in at least one sample were considered detected. The threshold value was found by inspection from the distribution plots of \log_2 expression levels. Detected probes were sorted according to their q value, which is the smallest false discovery rate (FDR) at which the probe is called significant. FDR was evaluated using significance analysis of microarrays and its implementation

in the official statistical package sam¹⁹. To avoid false positive calls due to spuriously small variances, the percentile of standard deviation values used for the exchangeability factor s_0 in the regularized t -statistic was set to 50. We combined the LESC and CEC samples into one group of four samples, and looked for differentially expressed genes between this group and SENCs samples. The top 100 significant genes in this comparison are presented in Extended Data Fig. 2. All genes in this figure are significant at the FDR level of 0.01 or less. A heatmap was created using in-house hierarchical clustering software, and colours qualitatively correspond to fold changes.

RNA-seq and hierarchical cluster analysis. Total RNA was purified by a Picopure RNA isolation kit (Life Technology). RNA-seq was performed as described previously²⁰. In brief, 600 ng of total RNA was first converted to cDNA by superscript III first strand synthesis kit with primer Biotin-B-T. The cDNA was purified by NucleoSpin Gel and PCR Clean-Up Kit column (Clontech) to remove free primers and enzyme. Then terminal transferase (NEB) was applied to block the terminal of a cDNA 3' end. Streptavidin-coated magnetic beads (Life Technology) were further applied to isolate cDNAs. After RNA degradation by sodium hydroxide, second-strand cDNA was synthesized by random priming with primer A-N8. The second strand cDNA was eluted from beads by heat denaturing. The cDNA was then used as template to construct libraries by amplifies with barcode primers and primer PB. The sequencing was done on Hiseq 2000 system.

Hierarchical cluster analysis was performed with cluster and Java TreeView²¹. The raw data were first filtered using default parameters provided by the program Cluster. The filtered data were further adjusted by log transformation, genes and arrays were centred by median, and then both gene and array were hierarchically clustered with euclidean and average linkage. The hierarchical trees and gene matrix were visualized and generated by Java Treeview.

Lentiviral RNA interference and PAX6 transduction. Lentiviral shRNAs targeting PAX6, WNT7A and FZD5 genes were either cloned into pLKO.1 plasmid between Age I and EcoR I or purchased directly from Sigma. shRNAs targeting sequences for gene-specific knockdowns were as follows: PAX6, CGTCCATCTT TGCTTGGGAAA and AGTTTGAGAGAACCCATTATC; WNT7A, CGTGCT CAAGGACAAGTACAA and GCGTTCACCTACGCCATCAT; FZD5, CGCG AGCCCTTCGTGCCATT and TCCTAAGGTTGGCGTTGTAAT. We used a lentiviral pLKO.1-puro Non-Target shRNA control plasmid encoding a shRNA that did not target any known genes from any species as a negative control in all gene knockdown experiments (Sigma).

Lentiviral shRNA particles were prepared according to a previous described protocol²². In brief, replication-incompetent lentiviral particles were packaged in 293T cells by co-transfection of shRNA constructs with packaging mix (pCMV-dR8.2 and pCMV-VSVG at a 9:1 ratio). Virus was collected two times at 48 h and 72 h post transfection.

For transduction, PAX6a open reading frame (ORF) was PCR amplified from cDNAs purchased from Thermo Scientific (MHS6278-202756612) and inserted into pLenti CMV-GFP Puro vector between BamHI and BsrGI. PAX6b was generated by PCR-mediated point mutation strategy with primers PAX6 InF and PAX6 InR (Extended Data Table 1; top). For GFP labelling, pLenti CMV-GFP Hygro (656-4) purchased from Addgene was used. The lentiviral particles were packaged by co-transfection with packaging plasmids psPax2 and pMD2.G.

For lentiviral infection, cells were infected for 16–20 h with fresh media containing individual virus and polybrene at a final concentration of $8 \mu\text{g ml}^{-1}$. The infected cells were further selected by $2 \mu\text{g ml}^{-1}$ puromycin for 48 h or $200 \mu\text{g ml}^{-1}$ hygromycin for 72 h.

Western blot analysis and co-immunoprecipitation. For western blot analysis, cells were washed once with PBS and then collected in cell lysis buffer (50 mM Tris-HCl, pH 6.8; 2% SDS; 10% Glycerol; 100 mM DTT). Protein concentration was quantified by Nanodrop and Bromophenol blue was added to a final concentration of 0.1%, then 25 μg of total lysate was fractionated on a 4–12% NUPAGE gel (Life Technology). Proteins were transferred to a nitrocellulose membrane at 100 V for 1 h. The membrane was blocked with 5% milk and probed with relevant antibodies and mouse anti- β -actin monoclonal antibody (A5316, Sigma).

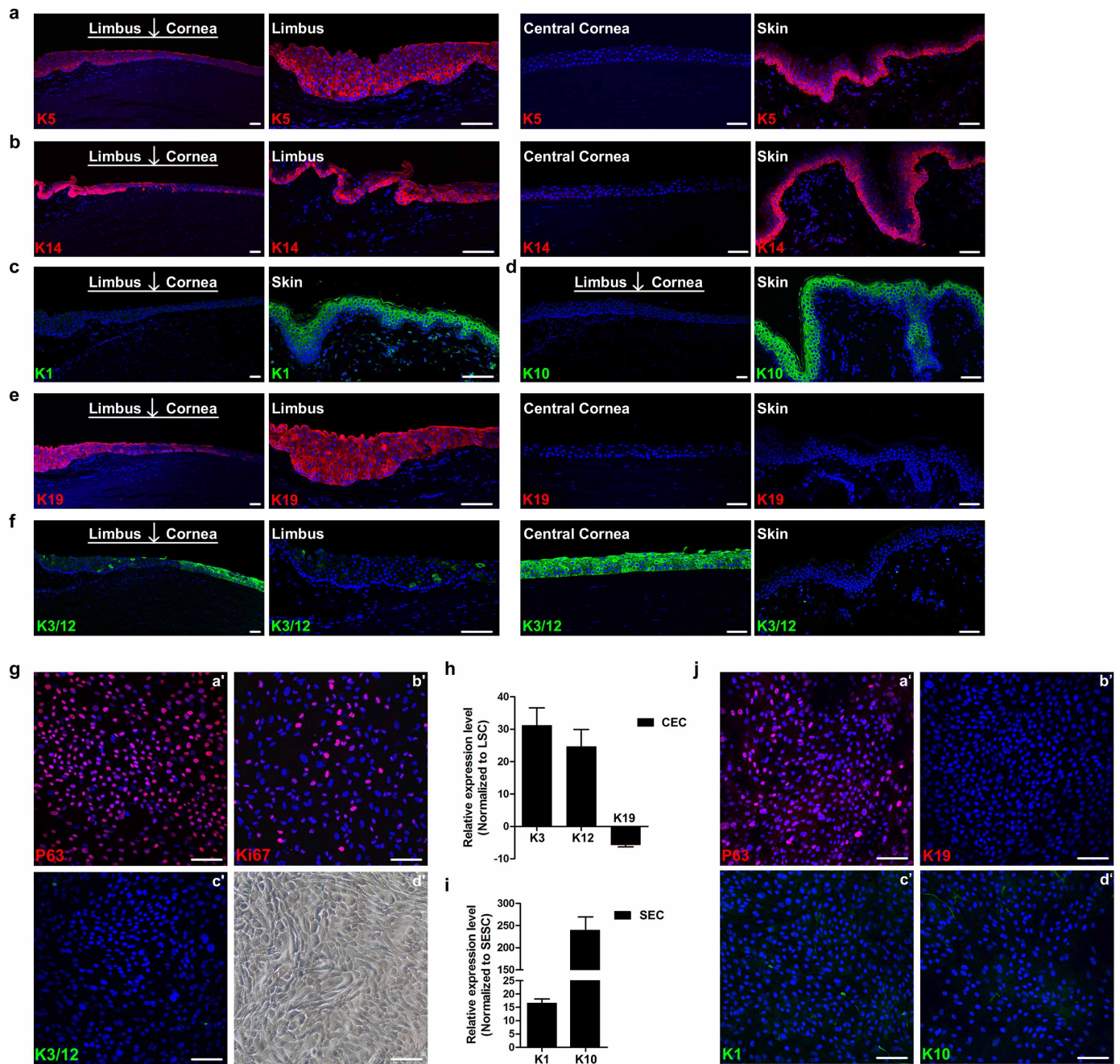
To detect interaction between FZD5 and WNT7A, a 10-cm dish of limbal stem cells at 90% confluence was collected; the cell pellet was resuspended in 700 μl of co-immunoprecipitation (Co-IP) buffer (10 mM Tris-HCl, pH 7.4, 100 mM NaCl, 2.5 mM MgCl_2 , 0.5% NP-40, $1 \times$ proteinase inhibitor) and incubated on ice for 20 min, then centrifuged at 13,000 r.p.m. at 4°C for 20 min. The 600 μl of supernatant were aliquoted into two pre-chilled Eppendorf tubes, 5 μg of rabbit anti-FZD5 monoclonal antibody (#5266, Cell Signaling) or WNT7A antibodies was added to each tube and incubated at 4°C overnight. Protein A/G magnetic beads (50 μl , Thermo Fisher) were added to each tube, and incubated at 4°C for 2 h, washed with a Co-IP buffer and eluted in $1 \times$ SDS sample buffer (Life Technology) at 70°C . The input and elutes were fractionated on 4–12% NUPAGE gel and blotted with FZD5 and WNT7A antibodies.

Cell transplantation. All animal studies were performed in full accordance with the Association for Research in Vision and Ophthalmology (ARVO) statement, Use of Animals in Ophthalmic and Vision Research, and approvals were obtained from Institutional Animal Care Committees.

New Zealand white rabbits (2.0 kg to 2.5 kg, male) were used in the study. Rabbits were anaesthetized with intramuscular injection of xylazine hydrochloride (2.5 mg ml^{-1}) and ketamine hydrochloride (37.5 mg ml^{-1}). To create a limbal stem cells deficiency model (Extended Data Fig. 7f), corneal and limbal epithelium was removed by 360-degree conjunctival peritomy and lamellar dissection to remove anterior scleral and corneal stromal tissues, 2 mm posterior from limbus towards the centre of the cornea. This dissection ensured removal of LSC and the entire corneal epithelium. Rabbit GFP-labelled LSCs (5×10^5), PAX6⁺ SENCs or shPAX6 LSCs cells were mixed with fibrin (25 mg ml^{-1}) and thrombin (25 U per ml) and seeded onto the exposed stromal bed of a recipient cornea and limbal area; the surface was then covered by a human amniotic membrane (Bio-tissue), which is secured with 10.0 VICRYL sutures (Ethicon) (Extended Data Table; bottom). As a negative control, only amniotic membrane was applied to the denuded cornea. Antibiotics (levofloxacin) and steroids (betamethasone) were applied to both eyes

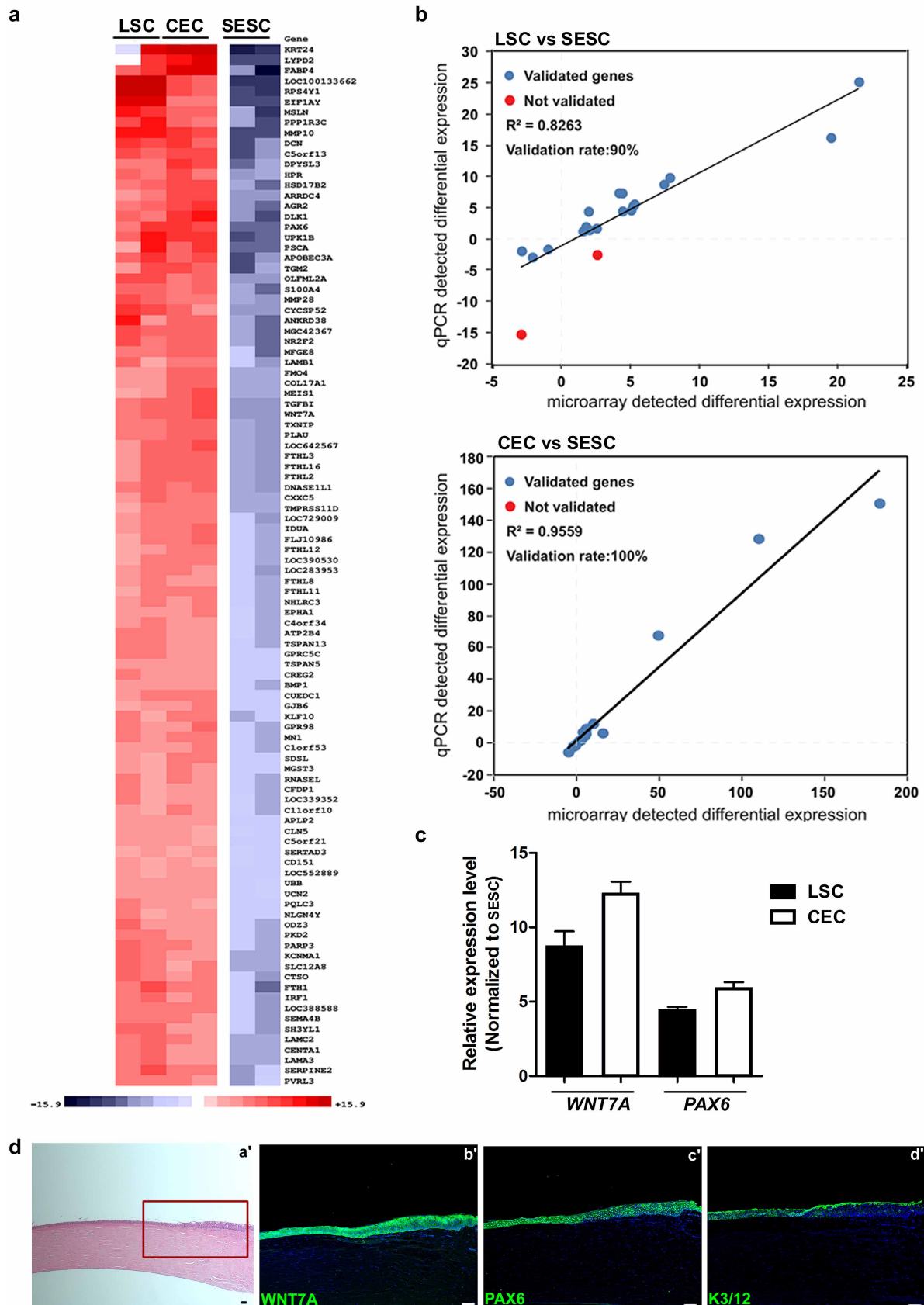
immediately after the cell transplant procedures, and were administered three times a day for 2 weeks. Animals were randomly assigned into each experimental group. The investigator who performed cell transplantation was blind to the identity of cells used. Another investigator carried out assessment of the effect of corneal epithelial repair in rabbits and was again blind to the identity of cells used in the transplantation. For analysis, we exclude only animals that died of post-operative complications such as infection, as they did not reach the end point for assessment of cell transplantation effect; this criterion is pre-established.

19. Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA* **98**, 5116–5121 (2001).
20. Fox-Walsh, K., Davis-Turak, J., Zhou, Y., Li, H. & Fu, X. D. A multiplex RNA-seq strategy to profile poly(A⁺) RNA: application to analysis of transcription response and 3' end formation. *Genomics* **98**, 266–271 (2011).
21. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* **95**, 14863–14868 (1998).
22. Xue, Y. *et al.* Direct conversion of fibroblasts to neurons by reprogramming PTB-regulated microRNA circuits. *Cell* **152**, 82–95 (2013).



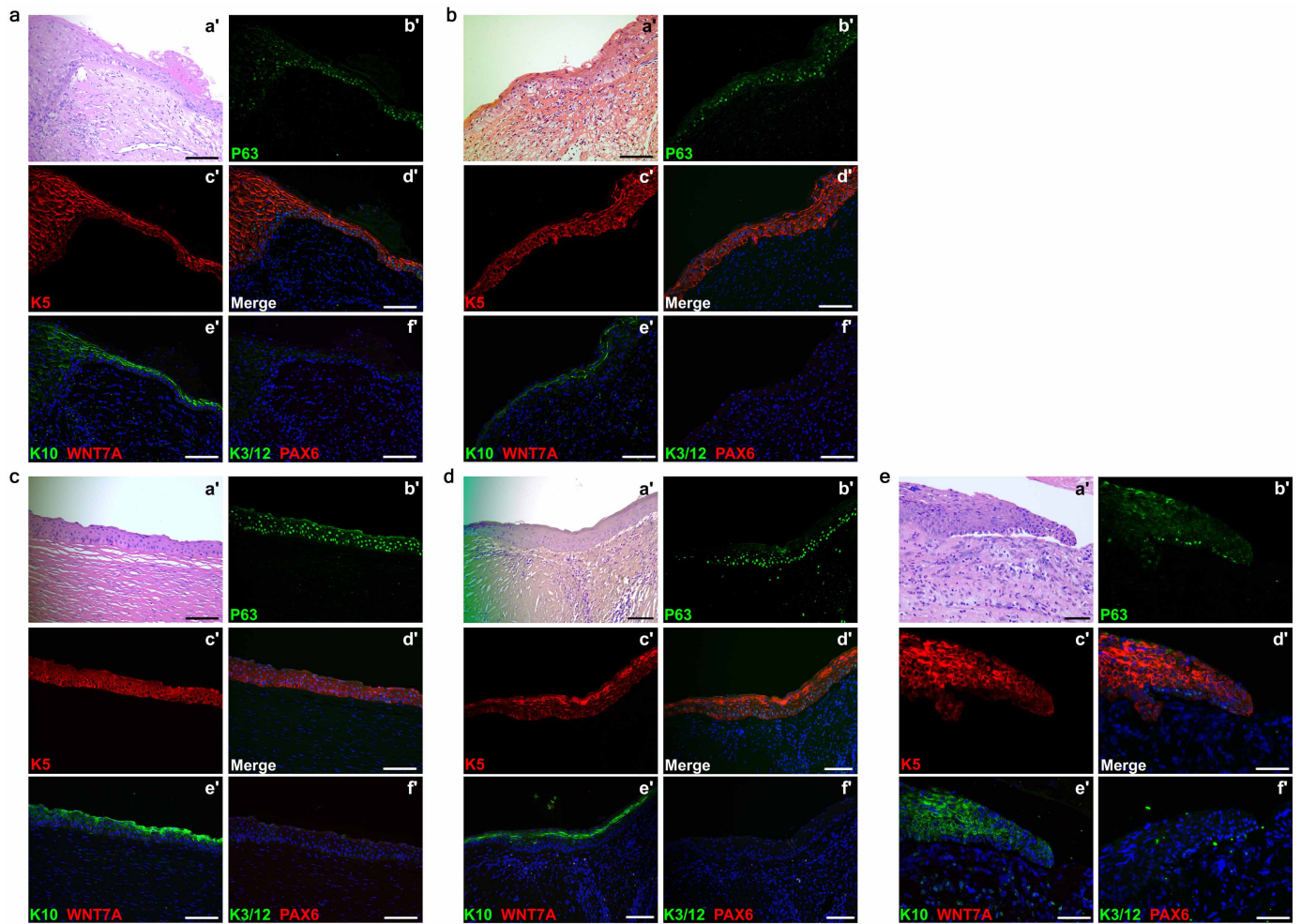
Extended Data Figure 1 | Keratin expression profiles and cell cultures, and three-dimensional differentiation of LSCs and SESC. **a–f**, Keratin expression profiles in human limbus, cornea and skin epidermis. **a**, **b**, Peripheral cornea–limbus junction and skin tissues showing K5⁺ (a) and K14 (b) expression in the basal cell layer of limbus and skin, and their absence in central corneal epithelium. **c**, **d**, Skin epidermis showing positive K1 (c) and K10 (d) expression and their absence in cornea and limbus. **e**, **f**, Peripheral cornea–limbus junction showing K19⁺ staining in limbus but not in central corneal epithelium and skin (e), and K3/12⁺ staining only in cornea and not in limbus and skin (f). **g–j**, Cultured LSCs with stem and progenitor cell characteristics, and SESC characteristics at passage 12 and validation of a

three-dimensional differentiation system. **g**, Immunofluorescence staining of LSCs showing positive stem cell signals of p63 (a') and Ki67 (b') and negative differentiated CEC signals, K3/12 (c'), phase contrast photograph (d'). **h**, qPCR analysis showing K3/12 upregulation and K19 downregulation in CECs from a three-dimensional differentiation assay compared with LSCs. **i**, K1 and K10 upregulation in SECs from three-dimensional differentiation assay compared with SESC (c), all $n = 3$, $P < 0.01$. **j**, Immunofluorescence staining of cultured SESC showing positive p63 (a') and negative signals for limbus stem cell marker, K19 (b') and mature skin epithelium markers K1 and K10 (c', d'). Scale bars, 100 μm . Data shown as means \pm s.d.



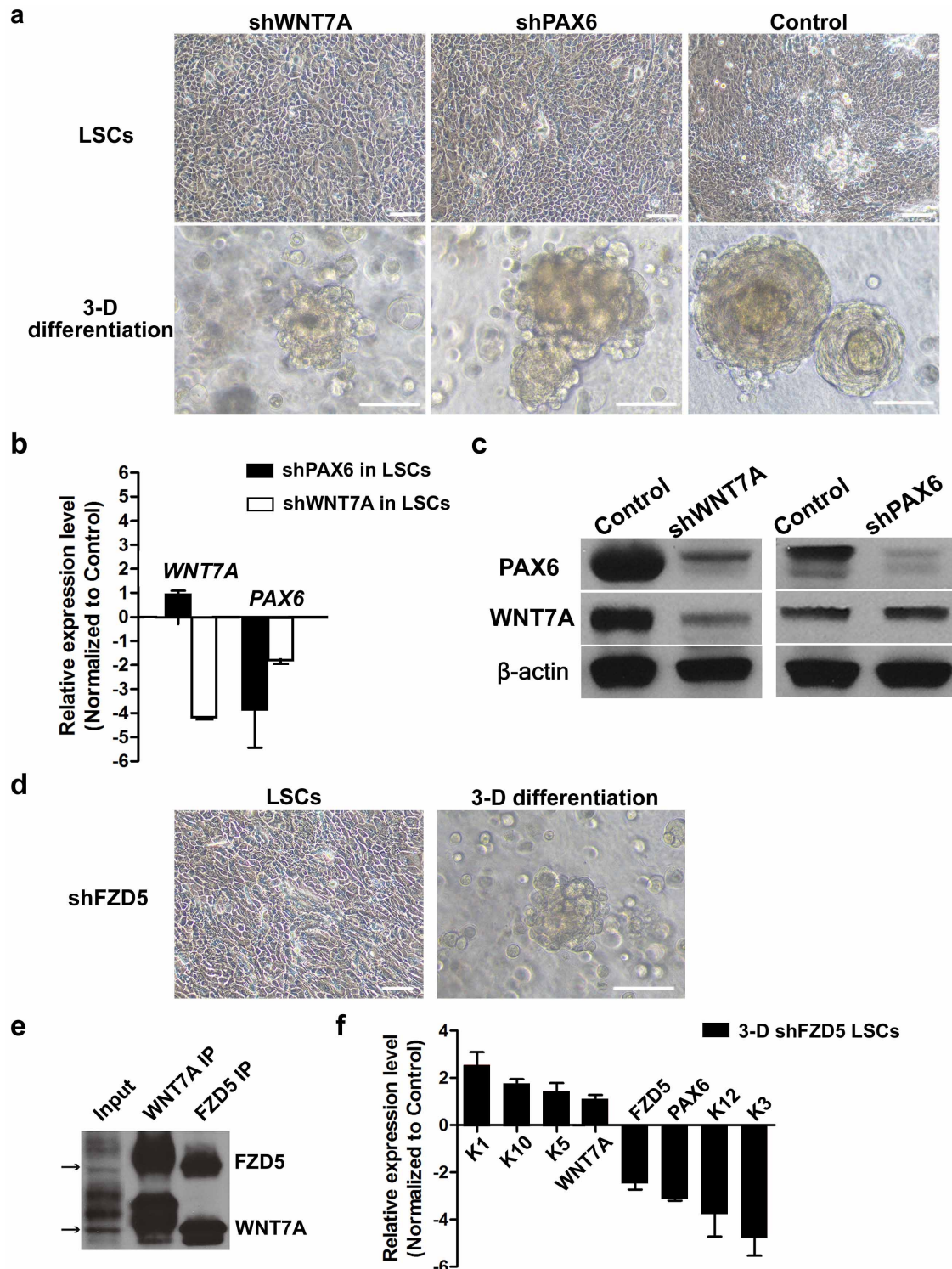
Extended Data Figure 2 | Gene expression profiling and immunohistological analysis. **a–c**, Genome-wide gene expression microarray of LSCs, CECs and SESC. **a**, The top 100 significant genes in a comparison of LSCs and CECs to SESC. **b**, Validation of the microarray data with qPCR analysis showing a strong correlation. **c**, qPCR analysis of *WNT7A* and *PAX6*

expression in LSCs and CECs compared to SESC, all $n = 3$, $P < 0.05$. **d**, Expression of *WNT7A* and *PAX6* in cornea and limbus of a one-year old human infant. H&E stain (**a'**), boxed area was shown in serial sections (**b'–d'**) with immunofluorescence staining of *WNT7A* (**b'**), *PAX6* (**c'**) and *K3/12* (**d'**). All scale bars, 100 μm . Data shown as means \pm s.d.



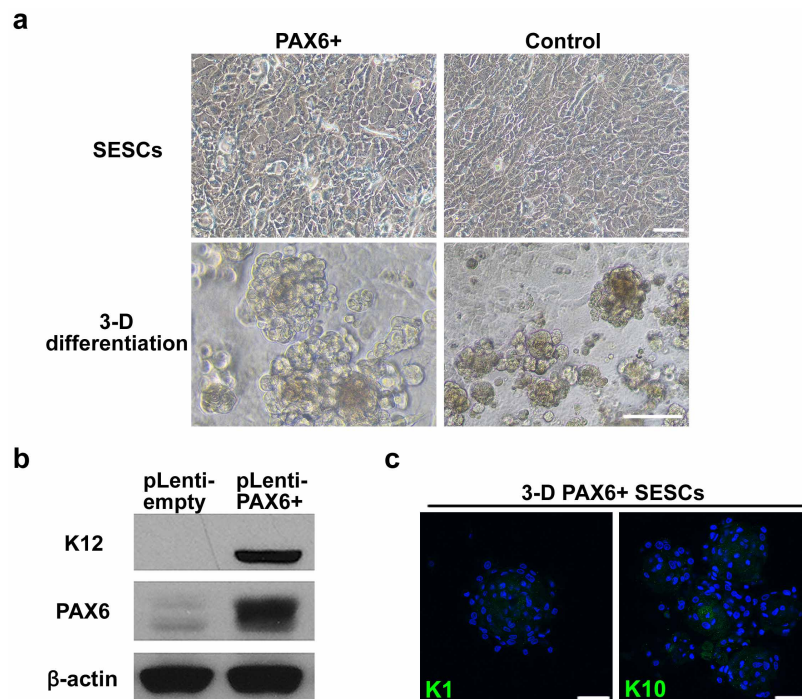
Extended Data Figure 3 | Appearance of skin epidermal markers with loss of corneal markers in human corneal diseases. Appearance of skin epidermal markers p63, K5 and K10 with loss of corneal marker K3/12, PAX6 and WNT7A in cornea of patients with Stevens–Johnson syndrome (a, b), ocular pemphigoid (c), trauma injury (d) and alkaline burn (e). For all images, H&E

staining was carried out on the lesion of corneal epithelial squamous metaplasia (a'). b'–f', the same region of lesion in serial sections showing increased p63 (b', d') and K5 (c', d') and K10 (e') in the suprabasal layer, no WNT7A (e'), K3/12 or PAX6 could be detected in the area (f'). Scale bars, 100 μ m.



Extended Data Figure 4 | The effect of WNT7A and FZD5 on PAX6 expression in LSCs. **a–c,** The effect of WNT7A knockdown on PAX6 expression in LSCs. **a,** Phase contrast photographs showing effects of WNT7A and PAX6 knockdowns (shWNT7A and shPAX6) in LSCs and their three-dimensional differentiation spheres. **b,** qPCR analysis of gene expression changes of WNT7A and PAX6 in LSCs. WNT7A knockdown decreased PAX6 expression ($n = 3$, $P < 0.01$); no significant change in WNT7A expression in PAX6 knockdown. **c,** Validation of knockdown efficiency of WNT7A and PAX6 in LSCs by western blot analysis. **d–f,** WNT7A and FZD5 acted as the

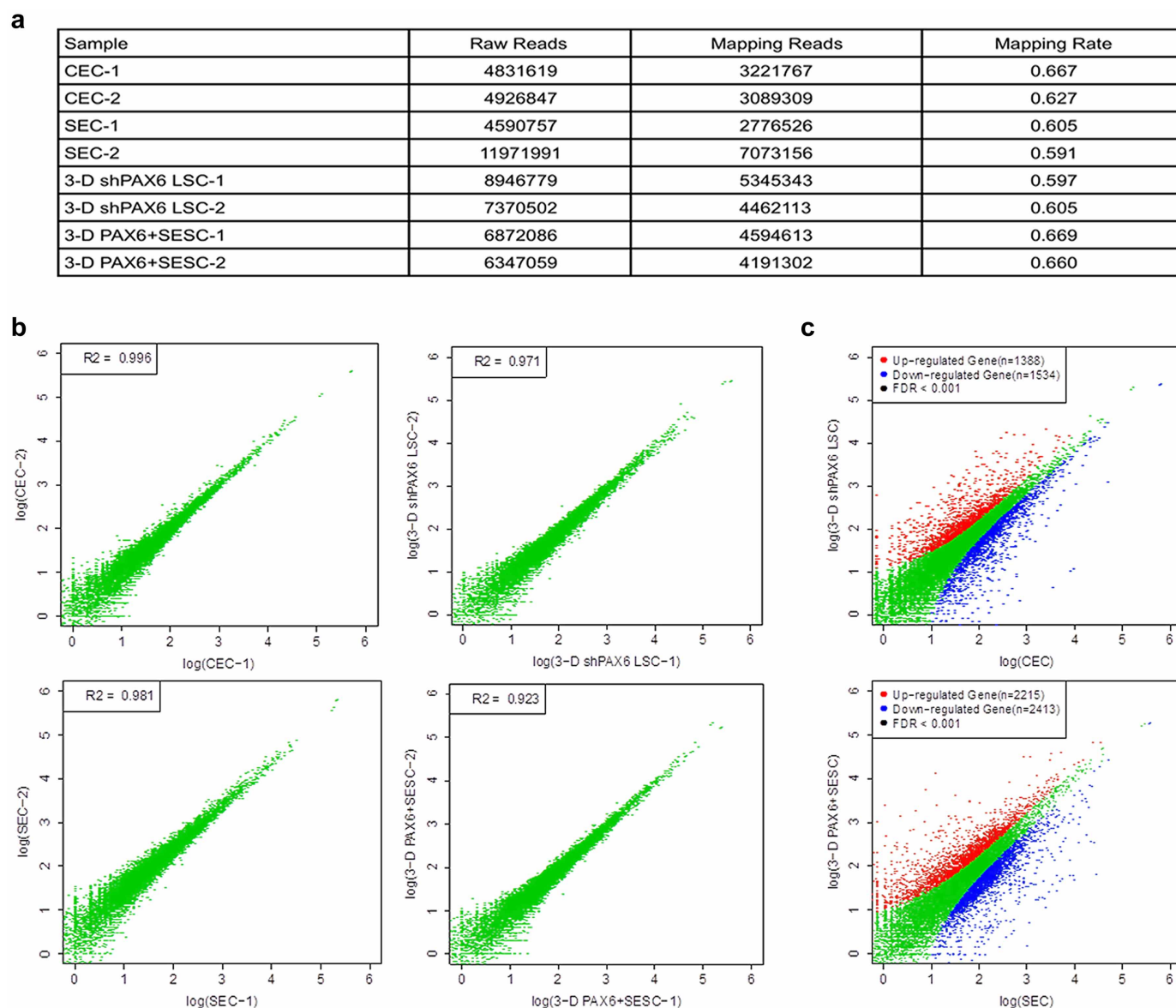
upstream regulators of PAX6 expression. **d,** Phase contrast photographs showing cell morphology of knockdown of FZD5 (shFZD5) in LSCs and three-dimensional differentiation spheres. **e,** Co-immunoprecipitation of WNT7A and FZD5 in LSCs. **f,** qPCR analysis of gene expression changes in corneal and skin epithelial markers in three-dimensional differentiated cells of LSCs with FZD5 knockdown (three-dimensional shFZD5 LSCs). FZD5 knockdown did not affect WNT7A expression; all others, $n = 3$, $P < 0.05$. Scale bars, 100 μ m. Data shown as means \pm s.d.



Extended Data Figure 5 | The effect of *PAX6* transduction in SENCs.

a, Phase contrast photographs of SENCs with *PAX6* transduction ($PAX6^{+}$) and three-dimensional differentiation spheres. **b**, Validation of K12 and *PAX6* expression in three-dimensional differentiation spheres by western blotting

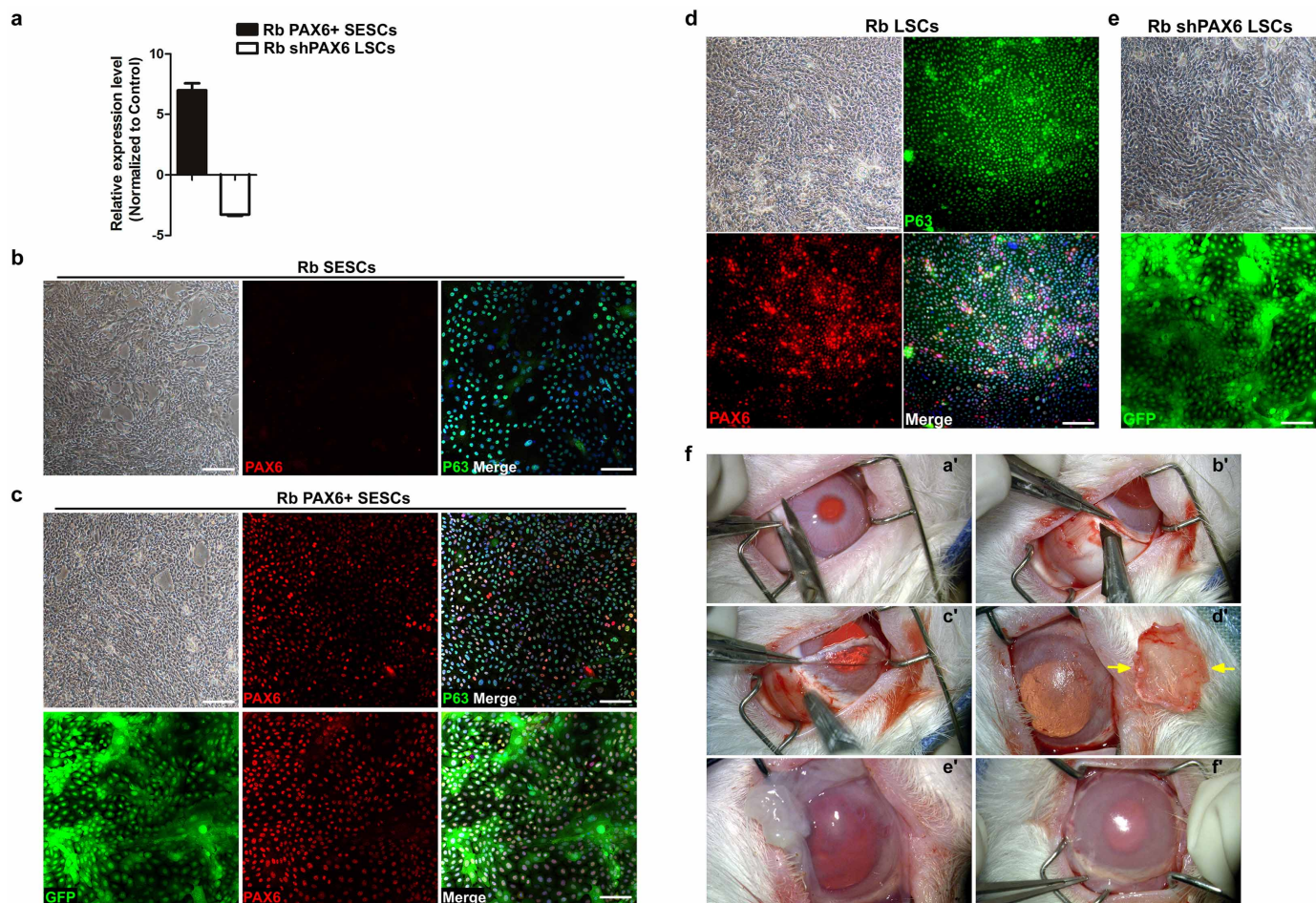
analysis. **c**, Loss of skin-specific keratins, K1 and K10 in three-dimensional differentiation of SENCs with *PAX6* transduction (three-dimensional $PAX6^{+}$ SENCs). Scale bars, 100 μ m.



Extended Data Figure 6 | Quantitative information from RNA-seq data.

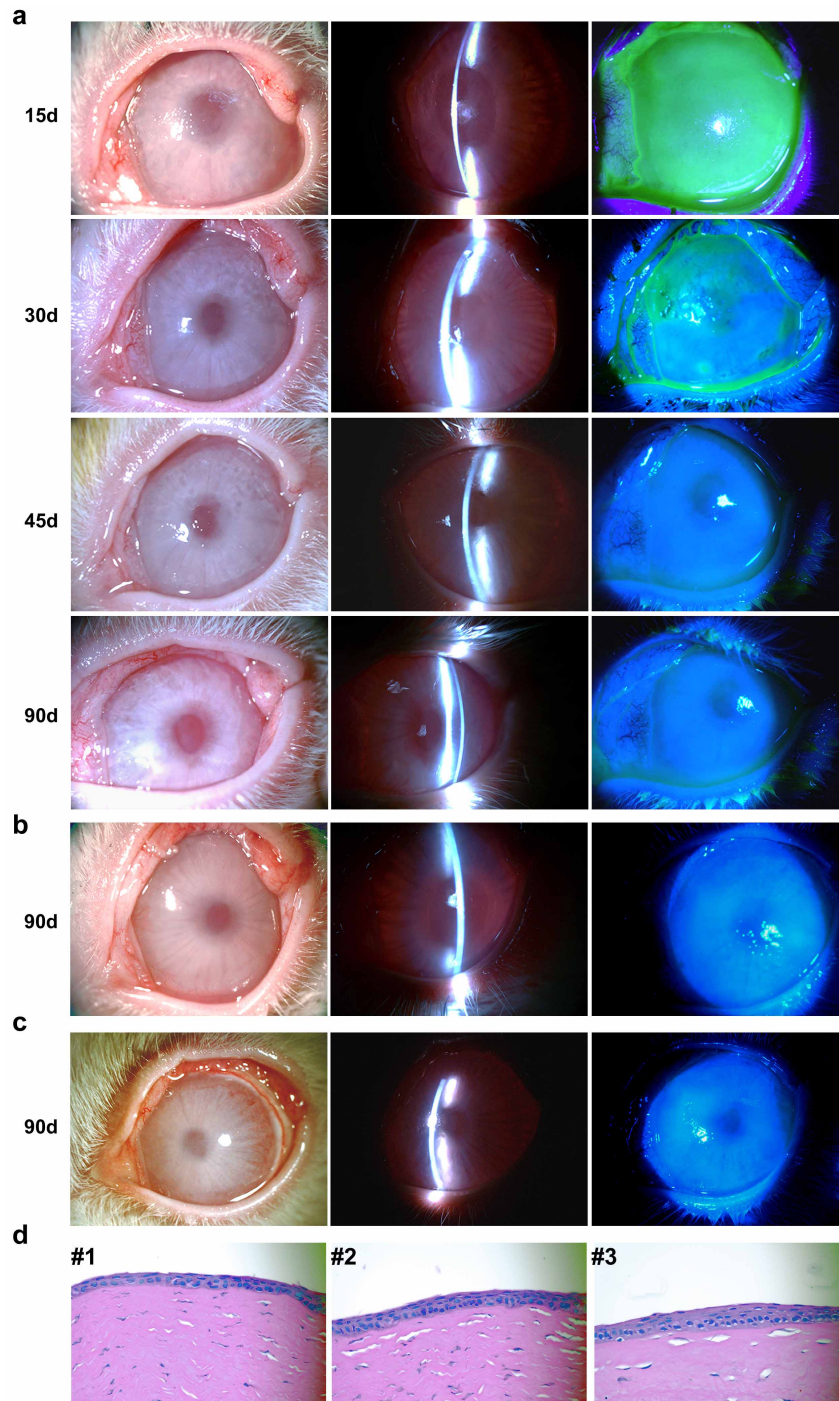
a, Statistical analysis of RNA-seq samples: raw reads, mapping reads and mapping rate of each sample are included. **b**, Pairwise comparisons of duplicated biological samples. **c**, The differences between SECs and three-dimensional PAX6⁺ SESC, CECs and three-dimensional shPAX6 LSCs, all FDR < 0.001. **a**, qPCR analysis of PAX6 expression in rabbit SESC with PAX6 transduction (rabbit (Rb) PAX6⁺ SESC) or LSCs with PAX6 knockdown

(Rb shPAX6 LSCs) (all $n = 3$, $P < 0.05$). We noticed some minor differences in the heatmap. These might result from some experimental variations, or it is possible that, although PAX6 expression is largely responsible for cell fate switch from SESC to CECs at gene expression and functional levels (as demonstrated in this study), this single transcription factor may not be sufficient to create cells that are completely identical to CECs.



Extended Data Figure 7 | Engineered expression of PAX6 and rabbit LSC deficiency model. **a–e**, Quantification and culture of engineered expression of PAX6 in rabbit SESC and PAX6 knockdown LSCs. **a**, qPCR analysis of PAX6 expression in rabbit SESC with PAX6 transduction (Rb PAX6⁺ SESC) or LSCs with PAX6 knockdown (Rb shPAX6 LSC) (all $n = 3$, $P < 0.05$). **b**, Rabbit SESC with positive staining of p63 and negative staining of PAX6. Left panel, phase contrast photograph. **c**, Top row, double immunofluorescence staining of PAX6 and p63 in rabbit SESC with PAX6 transduction. Top left panel, phase contrast photograph. Bottom row, rabbit PAX6⁺ SESC were further

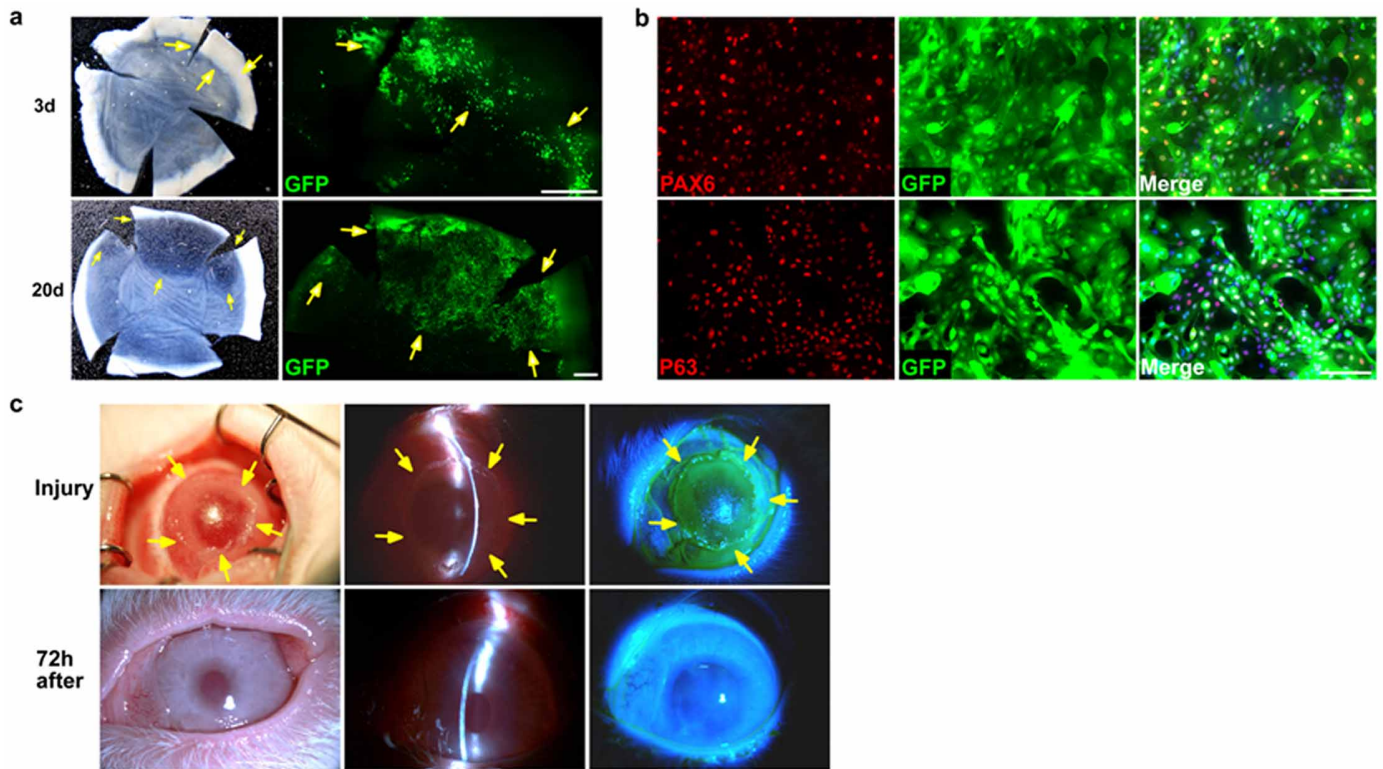
labelled with GFP for transplantation. **d**, Rabbit LSCs with positive staining of p63 and PAX6. Top left panel, phase contrast photograph. **e**, Culture of GFP-labelled rabbit LSCs with PAX6 knockdown. **f**, conjunctiva peritomy was performed and a circumferential strip of 2mm anterior limbal conjunctiva was removed (**a'**). Lamellar scleral and corneal dissection to completely remove LSCs and corneal epithelium along an anterior cornea stroma plane (**b'–d'**). Dissected cap is shown in (**d'**, arrows). The exposed cornea stroma bed was covered by human amniotic membrane (**e'**) and sutures (**f'**). ($n = 3$). Scale bars, 100 μ m. Data shown as means \pm s.d.



Extended Data Figure 8 | Cornea epithelium regeneration and repair by transplanted GFP-labelled PAX6⁺ SESC in a rabbit LSC deficiency model.

a, Time course of corneal epithelial defect repair. Fifteen days post transplantation, there was decreased cornea clarity with an entire corneal epithelial defect evidenced by fluorescein stain of cornea surface; 30 days post transplantation there was improved cornea clarity and reduced fluorescein staining of cornea epithelial defect; 45 and 90 days post transplantation there was restoration and maintenance of cornea clarity. **b, c,** Two other examples of regeneration and repair of rabbit corneal epithelial surface 90 days post

transplantation with GFP-labelled PAX6⁺ SESC showing complete repair and re-epithelization of corneal epithelial defects. Left panels, white light micrographs; middle panels, slit-lamp micrographs; right panels, fluorescein staining (note that bright spots on the corneal surface were due to camera light reflection, they were not epithelial defects) of corneal epithelium ($n = 5$). **d,** H&E stain of regeneration and repair of corneal epithelial surface in three separate rabbits 90 days post transplantation with GFP-labelled PAX6⁺ SESC showing intact corneal epithelium histology.



Extended Data Figure 9 | Corneal epithelial regeneration by transplantation in a rabbit LSC deficiency model. **a**, Time course of corneal epithelial regeneration and repair in a rabbit LSC deficiency model post transplantation with GFP-labelled PAX6⁺ SESC. Top panels, 3 day post transplantation. Left, light micrograph showing a hazy cornea; right, GFP⁺ donor cells at limbal region (arrows). Bottom panels, 20 days post transplantation. Left, light micrograph showing a cornea with partial clarity; right, GFP⁺ donor co-located in transparent areas (arrows). Scale bars, 1 mm. We observed that only the transplanted cells from the limbal region could survive, proliferate and regenerate cornea surface epithelium, suggesting that limbus contained the stem cell niche favourable for stem cell survival and growth. **b**, Culture and re-isolation of reprogrammed donor GFP-labelled

PAX6⁺ SESC epithelial cells from the limbal region of a rabbit recipient eye 90 days post transplantation with GFP-labelled PAX6⁺ SESC. Top panel, double immunofluorescence staining of PAX6 and GFP; bottom panel, double immunofluorescence staining of p63 and GFP in PAX6-transduced rabbit SESC. Scale bars, 100 μ m. **c**, Repair and recovery of a repeat cornea epithelium injury on a cornea transplanted with GFP-labelled PAX6⁺ SESC. Top panels, we iatrogenically scraped and removed donor-derived corneal epithelial cells and made a large corneal surface epithelium defect (arrows) 3 months post initial transplantation of PAX6⁺ SESC. Bottom panels, complete repair and recovery were observed within 72 h with healed epithelial defect ($n = 3$). Left panels, light micrographs; middle panels, slit-lamp micrographs; right panels, fluorescein staining.

Extended Data Table 1 | Primer sequences and rabbit transplantation results

Extended Data Table 1a. Primer sequences

Gene (Human)	Forward Primer	Reverse Primer
<i>CASZ1</i>	GTTCTACGGACAGAAGACCACG	TCTTGAAGCCGTCCTTGGCGTA
<i>FGFR3</i>	AGTGGAGCCTGGTCATGGAA	GGATGCTGCCAAACTTGTCTC
<i>FZD5</i>	TGGAACGCTTCCGCTATCCTGA	GGTCTCGTAGTGGATGTGGTTG
<i>GAPDH</i>	GAGTCAACGGATTTGGTCGT	GACAAGCTTCCCGTTCTCAG
<i>ID2</i>	TTGTCAGCCTGCATCACCAGAG	AGCCACACAGTGCTTTGCTGTC
<i>K1</i>	CAGCATCATTGCTGAGGTCAAGG	CATGTCTGCCAGCAGTGATCTG
<i>K3</i>	ACGTGACTACCAGGAGCTGATG	ATGCTGACAGCACTCGGACACT
<i>K5</i>	GCTGCCCTACATGAACAAGGTGG	ATGGAGAGGACCCTGAGGTGT
<i>K10</i>	CCTGCTTCAGATCGACAATGCC	ATCTCCAGGTGAGCCTTGGTCA
<i>K12</i>	AGCAGAATCGGAAGGACGCTGA	ACCTCGCTCTTGTGAGCTGAA
<i>K14</i>	TGCCGAGGAATGGTTCTTCACC	GCAGCTCAATCTCCAGGTTCTG
<i>K15</i>	AGGACTGACCTGGAGATGCAGA	TGCGTCCATCTCCACATTGACC
<i>K19</i>	AGCTAGAGGTGAAGATCCGCGA	GCAGGACAATCCTGGAGTTCTC
<i>MEIS1</i>	AAGCAGTTGGCACAAGACACGG	CTGCTCGGTTGAGCTGGTCTAT
<i>MMP9</i>	GCCACTACTGTGCCCTTTGAGTC	CCCTCAGAGAATCGCCAGTACT
<i>MMP10</i>	TCCAGGCTGTATGAAGGAGAGG	GGTAGGCATGAGCCAACTGTG
<i>NR2F2</i>	TGCACGTTGACTCAGCCGAGTA	AAGCACACTGAGACTTTTCCTGC
<i>NOTCH1</i>	GGTGAAGTCTCTGAGGAGATC	GGATTGCAGTCTGCCACGTTGA
<i>NOTCH3</i>	TACTGGTAGCCACTGTGAGCAG	CAGTTATCACCATTGTAGCCAGG
<i>ODZ3</i>	GGACAAGGCTATCACAGTGGAAC	TTCTGAGGGAGCCGTCATAACC
<i>PAX6</i>	TGTCCAACGGATGTGTGAGT	TTTCCAAGCAAAGATGGAC
<i>PDGFA</i>	CAGCGACTCCTGGAGATAGACT	CGATGCTTCTTCTCCGAATG
<i>PPARG</i>	AGCCTGCGAAAGCCTTTTGGTG	GGCTTCACATTCAGCAAACCTGG
<i>PRDM8</i>	CTGTGTCTTGAGCCATACTTCC	CCTTCTGAGGAACCAATTGCTGC
<i>TGFBI</i>	AGGACTGACGGAGACCCCAAC	TCCGCTAACCAGGATTTCATCAC
<i>WNT7A</i>	TGCCCCGACTCTCATGAAC	GTGTGGTCCAGCACGCTTTG
Gene (Rabbit)	Forward Primer	Reverse Primer
<i>GAPDH</i>	GCGAGATCCCGCCAACATCAAGT	AGGATGCGTTGCTGACAATC
<i>PAX6</i>	GTATTCTTGCTTCAGGTAGAT	GAGGCTCAAATGCGACTTCAGCT
Primers used for PAX6 transduction		
<i>PAX6 InF</i>	TTCCCGAATTCTGCAGACCCATGCAGATGCAAAAGTCCAAGTGCTGGACAATCAAACGTTGCCAACGGATGTG	
<i>PAX6 InR</i>	CACATCCGTTGGACACGTTTTGATTGTCCAGCACTTGGACTTTTGCATCTGCATGGGTCTGCAGAATTCGGGAA	

Extended Data Table 1b. Summary of rabbit transplantation results

GFP-labeled donor cells	Regeneration and re-epithelization	Rabbit number	
		Opaque and vascularized corneal surface	Died from systemic infection or unrelated complications
LSCs	3	0	0
PAX6+ SESC	5	0	2
shPAX6 LSCs	0	4	1

a, Primer sequences for human and rabbit genes used in this study. b, Corneal regeneration and re-epithelization were arrayed three months after transplantation.

BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2

Vaibhav Bhatia¹, Sonia I. Barroso¹, María L. García-Rubio¹, Emanuela Tumini¹, Emilia Herrera-Moyano¹ & Andrés Aguilera¹

Genome instability is central to ageing, cancer and other diseases. It is not only proteins involved in DNA replication or the DNA damage response (DDR) that are important for maintaining genome integrity: from yeast to higher eukaryotes, mutations in genes involved in pre-mRNA splicing and in the biogenesis and export of messenger ribonucleoprotein (mRNP) also induce DNA damage and genome instability. This instability is frequently mediated by R-loops formed by DNA–RNA hybrids and a displaced single-stranded DNA¹. Here we show that the human TREX-2 complex, which is involved in mRNP biogenesis and export, prevents genome instability as determined by the accumulation of γ -H2AX (Ser-139 phosphorylated histone H2AX) and 53BP1 foci and single-cell electrophoresis in cells depleted of the TREX-2 subunits PCID2, GANP and DSS1. We show that the BRCA2 repair factor, which binds to DSS1, also associates with PCID2 in the cell. The use of an enhanced green fluorescent protein-tagged hybrid-binding domain of RNase H1 and the S9.6 antibody did not detect R-loops in TREX-2-depleted cells, but did detect the accumulation of R-loops in BRCA2-depleted cells. The results indicate that R-loops are frequently formed in cells and that BRCA2 is required for their processing. This link between BRCA2 and RNA-mediated genome instability indicates that R-loops may be a chief source of replication stress and cancer-associated instability.

R-loops can negatively affect transcription elongation^{2,3}; they have also been involved in promoter-proximal pausing and termination of transcription^{4–6}. However, R-loops also mediate stalling of the replication fork as a source of genome instability^{7–9}, and evidence suggests that they form at common fragile sites¹⁰. Factors such as senataxin or RNase H1 (RNH1) are implicated in R-loop dissolution¹, but the mechanisms used by the cellular machinery to resolve R-loops are poorly understood. Several mRNP biogenesis factors prevent R-loop formation and transcription-associated genome instability, an example being the conserved eukaryotic THO complex involved in transcription elongation and mRNA processing and export^{2,11–13}. THSC (also known as TREX-2) is another well-conserved protein complex working in RNA export, preferentially located at the nuclear pore complex^{14–16} (Extended Data Fig. 1). In yeast it has a similar effect on genome instability to that of THO¹⁷, and in humans it is composed of PCID2, GANP, CENP and DSS1 (refs 14, 18). However, there is no information about the role of human TREX-2 in genome integrity.

We therefore analysed genome stability in HeLa cells that were depleted of PCID2, GANP and DSS1 by short interfering RNA (siRNA) (Extended Data Fig. 2); for this purpose we determined γ -H2AX and 53BP1 foci (Fig. 1a). These foci increased after the depletion of both PCID2 and DSS1. Consistent with DNA break accumulation, alkaline single-cell electrophoresis revealed an increase in tail moment (Fig. 1b). Depletion of the SAC3 homologue in human TREX-2 GANP also induced genomic instability, but to a smaller extent (Extended Data Fig. 3a, b). Because one cause of genome instability is transcription–replication collisions, which are augmented by R-loops^{7,9}, we used DNA combing to measure the effect of TREX-2 depletion on the progress of the replication fork. The velocity of the replication fork in cells depleted of PCID2 and GANP was slightly faster than in control cells, and track

lengths were longer (Fig. 1c and Extended Data Fig. 3c), a phenomenon previously observed in cells depleted of other RNA-processing factors¹¹. In contrast, depletion of DSS1 led to slower-moving forks; the inter-origin distance was longer, implying a lower density of active replicons, especially in PCID2-depleted cells (Fig. 1c). These results suggest that the causes of genome instability involved different alterations in the replication profile.

In addition to PCID2, DSS1 binds and stabilizes BRCA2, a double-strand break repair factor¹⁹. We therefore determined whether PCID2 and BRCA2 interact *in vivo* by using a proximity ligation assay, which detects cellular protein–protein interactions *in situ*²⁰. We found a close

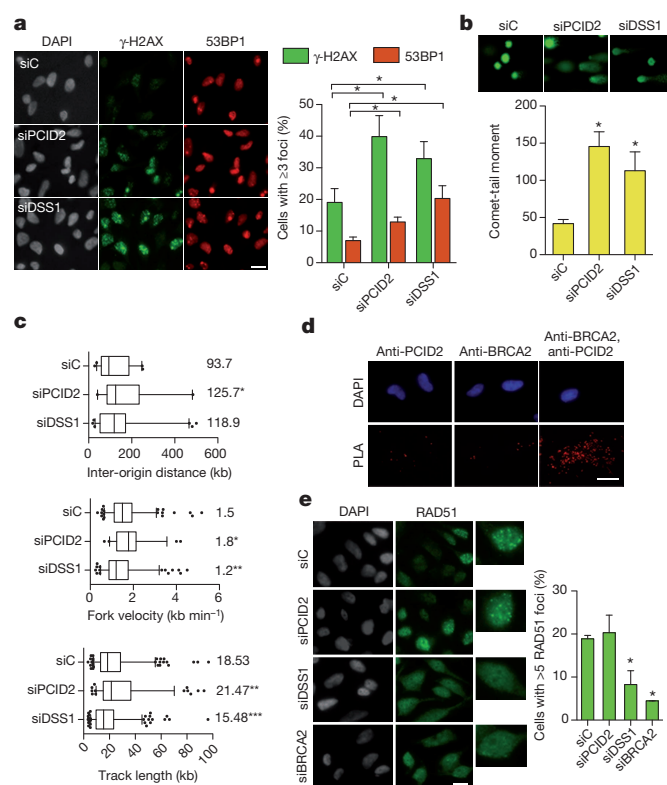


Figure 1 | TREX-2 associates with BRCA2 and affects genome integrity.

a, γ -H2AX and 53BP1 foci in siRNA-transfected cells ($n = 5$). Means and s.e.m. are plotted. DAPI, 4',6-diamidino-2-phenylindole. **b**, Single-cell electrophoresis in siRNA-treated cells. Relative comet-tail moments are plotted ($n = 3$) as means and s.e.m.. **c**, DNA-combing analysis of replication. The median, 25th to 75th centile range (boxes) and 5th to 95th centile range (whiskers) are plotted ($n = 3$). **d**, Proximity ligation assay showing specific interactions of endogenous BRCA2 and PCID2 (red spots) ($n = 3$). **e**, RAD51 foci in siRNA-transfected HeLa cells ($n = 5$). siC is the non-targeted control siRNA. Means and s.e.m. are plotted. * $P \leq 0.05$; ** $P < 0.01$ (two-tailed Student's *t*-test (**a**, **b**, **e**) or Mann–Whitney test (**c**)). Scale bars shown in all figures are 20 μ m.

association of the two proteins in the cells (Fig. 1d and Extended Data Fig. 4). Next we assessed whether or not the accumulation of DNA damage in PCID2-depleted cells was due to a defect in RAD51 loading. As expected, depletion of BRCA2 (as a control) and of DSS1, but not of PCID2, decreased the number of RAD51 foci (Fig. 1e), indicating that genomic instability in PCID2-depleted cells is not related to a functional defect of RAD51.

Next we examined whether DNA–RNA hybrids accumulated in the absence of PCID2, DSS1 and BRCA2, as a source of instability. For this, a fusion protein of the 52-residue DNA–RNA hybrid-binding (HB) domain of RNH1 and enhanced green fluorescent protein (EGFP) was constructed, hereafter referred to as HB–GFP (Fig. 2a). We confirmed that HB–GFP was able to detect RNA–DNA hybrids *in vivo* by several methods. Confocal microscopy of HeLa cells transfected with vectors encoding HB–GFP or EGFP revealed that, unlike EGFP, HB–GFP localized preferentially in the nucleus (Fig. 2b). When cells were permeabilized with detergent, EGFP was totally washed out, whereas HB–GFP was retained in the nucleus (Fig. 2c). Furthermore, size-exclusion chromatography of HB–GFP expressed in HEK293 cells showed a fraction of the cellular protein pool to be part of a multimeric complex, and reversible crosslink immunoprecipitation revealed HB–GFP interaction with histone H3 and single-stranded DNA (ssDNA)-binding replication protein A, consistent with an association of HB–GFP with chromatin (Extended Data Fig. 5a, b). Immunostaining with anti-GFP in HB–GFP-transfected HeLa cells and the anti-DNA–RNA S9.6 antibody in

HeLa cells showed nuclear localization and an intense staining in the nucleolar region (Fig. 2d), consistent with the predicted accumulation of R-loops in the ribosomal DNA region²¹. We confirmed this by chromatin immunoprecipitation (ChIP), which showed increased HB–GFP recruitment at the ribosomal DNA after depletion of TOP1 (Extended Data Fig. 5c), as expected. A pool of cells immunostained with anti-GFP and S9.6 also showed an accumulation of signal near the nuclear periphery (Fig. 2e). Further, HB–GFP crosslinked with formaldehyde to pelleted chromatin could be released in the supernatant of cell lysates by treatment with RNH1 (Extended Data Fig. 5d). Cells depleted of the THO subunit THOC1, which are known to accumulate DNA–RNA hybrids¹¹, were used as a positive control with the expected positive results. Taken together, these results confirm that HB–GFP efficiently detects R-loops *in vivo*.

We then used flow cytometry to show quantitatively that, after permeabilization, HB–GFP retention led to a significant number of fluorescence-positive cells, unlike EGFP, which was completely washed out. Overexpression of active RNH1, which competes with HB–GFP, decreased the number of HB–GFP-positive cells (Fig. 2f and Extended Data Fig. 5e). We therefore used this assay to quantify DNA–RNA hybrids as the number of HB–GFP-positive cells without RNH1 overexpression relative to the number with RNH1 overexpression (Fig. 2g). DNA–RNA hybrids increased slightly in the siTOP1 (*TOP1* siRNA-treated) control (Extended Data Fig. 5f) and did not increase in PCID2-depleted cells, but strongly increased in BRCA2-depleted and DSS1-depleted cells, in a similar manner to the siTHOC1 control (Fig. 2g). BRCA1-depleted cells also increased HB–GFP retention, but to a smaller extent.

Next we analysed DNA–RNA hybrid accumulation at various genes at the molecular level. ChIP with HB–GFP showed a significant accumulation of hybrids in BRCA2-depleted cells in the three genes assayed, whereas RNH1 overexpression decreased the signal (Fig. 3a and Extended Data Fig. 6). The siTHOC1 control behaved consistently. DNA–RNA immunoprecipitation (DRIP) using the S9.6 antibody²² showed that BRCA2 depletion increased DNA–RNA hybrids at four actively transcribed genes, whereas the siTHOC1 cells showed a heterogeneous profile (Fig. 3b) consistent with this being dependent on parameters such as transcription levels or gene length in yeast¹³. R-loop accumulation did not increase in PCID2-depleted cells (Fig. 3a, b), but co-depletion of BRCA2 and PCID2 slightly increased HB–GFP recruitment (Extended Data Fig. 7), consistent with the notion that both TREX-2 and BRCA2 could, at least in part, cooperate to prevent R-loop formation. Because BRCA2 functions at double-strand breaks by recruiting Rad51, and at RFs as part of the Fanconi anaemia pathway, we examined whether R-loops accumulated after depletion of Rad51 or BRCA1 to obtain a further insight into the role of BRCA2 in R-loops. BRCA1-depleted cells, but not RAD51-depleted cells, accumulated R-loops as detected by DRIP (Fig. 3c), suggesting that BRCA2 prevents R-loops independently of the Rad51 homologous recombination function. Indeed, R-loops were accumulated in non-replicating and replicating BRCA2-depleted cells as detected by immunofluorescence, in the latter case at a substantially higher level (Fig. 4a). Spontaneous recruitment of BRCA2 at different genes was clearly decreased by RNH1 overexpression as detected by ChIP in four genes analysed (Fig. 4b), indicating that BRCA2 is recruited to DNA–RNA hybrid regions.

We then speculated that expression of the inactive HB–GFP should bind and stabilize DNA–RNA hybrids, consequently increasing genome instability by strengthening a putative block of the replication fork. As expected, cells depleted of BRCA2, DSS1 and THOC1, but not of PCID2, showed elevated levels of γ -H2AX foci in HB–GFP-expressing cells (Fig. 4c). Chromosomal abnormalities are a hallmark of cancer and BRCA2-depleted cells, and we found that expression of HB–GFP increased chromosome breaks and decreased spontaneous sister chromatid exchanges (Extended Data Fig. 8a, b), as shown for BRCA2-deficient mouse fibroblasts²³ and Fanconi anaemia cells²⁴, even though BRCA2-deficient tumour cells show increased sister chromatid exchange²⁴. In addition, the genomic instability in BRCA2-depleted and HB–GFP-expressing retinal pigment

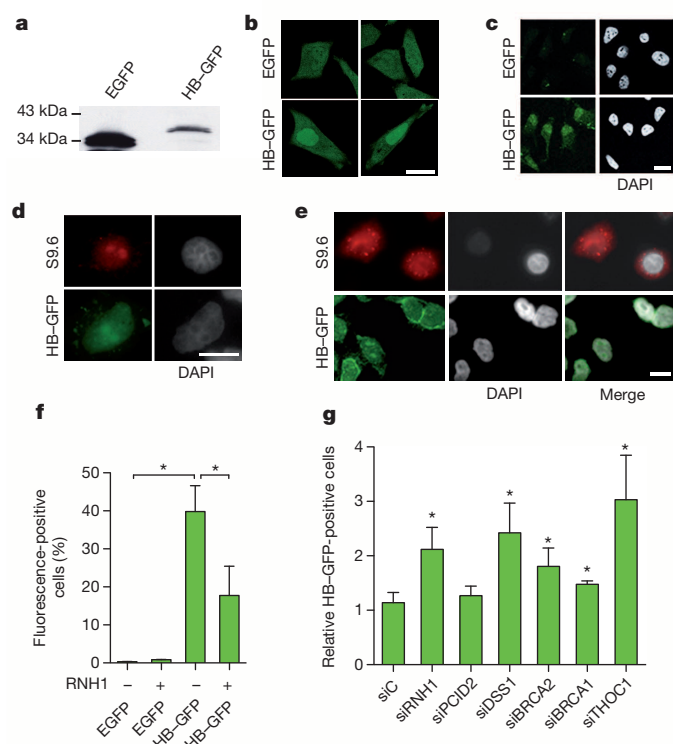


Figure 2 | Detection of DNA–RNA hybrids in mammalian cells *in vivo*. **a**, Immunoblot showing HB–GFP and EGFP expression in HeLa cells. **b**, **c**, Fluorescence confocal-microscopy of HeLa cells expressing EGFP and HB–GFP without (**b**) and with (**c**) detergent permeabilization. **d**, Immunofluorescence with S9.6 or anti-GFP antibodies in HeLa or HB–GFP-expressing HeLa cells, respectively, after formaldehyde fixation and permeabilization. **e**, Immunofluorescence in pre-permeabilized and fixed HB–GFP-expressing HeLa cells. Z-stacks highlighting the nuclear membrane signal for S9.6 and HB–GFP are shown ($n = 3$). **f**, DNA–RNA hybrid quantification by FACS. Means and s.e.m. of GFP-positive or HB–GFP-positive cells are plotted ($n = 3$) (see also Extended Data Fig. 5e). **g**, FACS quantification of HB–GFP-positive cells. The ratio of RNH1-untransfected to RNH1-transfected cells is shown ($n = 5$). Means and s.e.m. are plotted. * $P \leq 0.05$ (two-tailed Student's *t*-test).

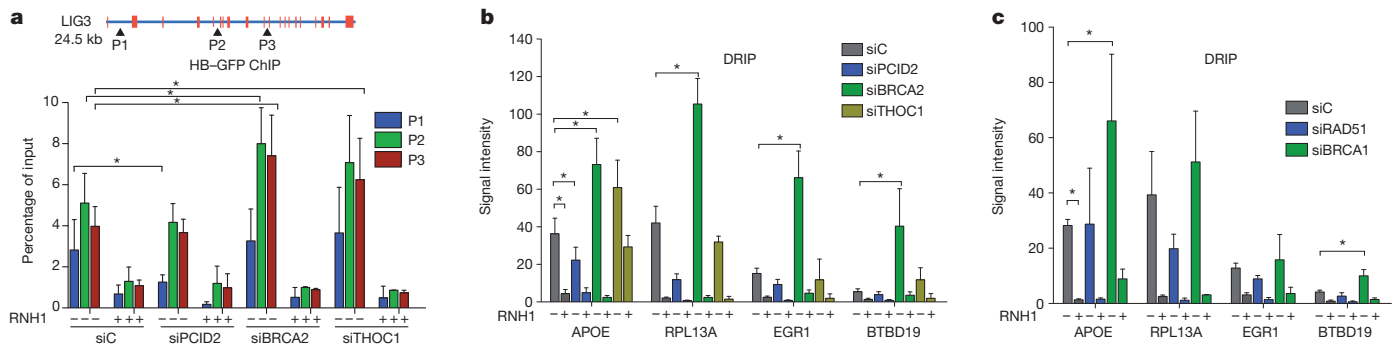


Figure 3 | DNA-RNA hybrids accumulate in actively transcribed genes in BRCA2 and THOC1-depleted cells. **a**, ChIP analysis of LIG3 in siRNA-transfected HeLa cells expressing HB-GFP. Percentage of input is plotted ($n = 3$). Positions of the amplicons used for LIG3 are shown. **b, c**, DRIP with

the S9.6 antibody. Signal intensity, relative to the SNRPN negative control region, is plotted ($n = 3$). Means and s.e.m. are plotted in all panels. $*P \leq 0.05$ (Mann-Whitney test).

epithelial (RPE-1) cells was correlated with increased adhesion-independent proliferation, a hallmark of cell transformation (Extended Data Fig. 8c), which is consistent with a potential role of R-loops as an intermediate in cancer risk. This is consistent with an increase in DNA-RNA hybrids in BRCA2-deficient cells enhancing HB-GFP binding and decreasing the efficiency of repair. Finally, we extended our study to the BRCA2-defective pancreatic adenocarcinoma-derived CAPAN-1 cells. These cells show pan-nuclear staining of γ -H2AX, as a signal of replication stress; this disappeared after RNH1 overexpression (Fig. 4d). Any remnant γ -H2AX observed in CAPAN-1 cells overexpressing RNH1 was consistently

seen in proximity to the nuclear periphery, in agreement with data in Fig. 2e.

Taken together, our results demonstrate that R-loops accumulate at high levels in BRCA2-depleted cells. This accumulation is seen in both replicating and non-replicating cells. Because TREX-2/PCID2 works on RNA export close to the nuclear pore complex^{14,16}, it is possible that the topological constraints resulting from the physical tethering of transcribed genes to the nuclear pore complex²⁵ promote R-loop accumulation and replication stress. Depletion of TREX-2 could loosen the connection of the transcribed site to the nuclear pore complex¹⁵, minimally contributing to R-loop accumulation in cells. It may also be that, in the absence of TREX-2, other RNA-binding factors help to prevent R-loop accumulation to a level detectable in the conditions tested here. We therefore propose that TREX-2 helps recruit BRCA2 to the proximity of transcribed regions, where BRCA2 binds to the branched structure formed by the displaced ssDNA and the DNA-RNA hybrids of naturally formed R-loops (Extended Data Fig. 9a). Such a structure may mimic the replicative intermediate to which BRCA2 is believed to bind *in vivo*^{26,27}. This could help expose the DNA-RNA hybrid for greater access by enzymes that remove the RNA chain, such as RNases (RNH1) or DNA-RNA helicases (senataxin). This is supported by the observations that BRCA2 is recruited to DNA-RNA hybrid regions and that HB-GFP expression enhances genome instability in normal cells and has a synergistic effect on BRCA2-depleted cells. In replicating chromatin, because BRCA2 and BRCA1 work in the Fanconi anaemia pathway, which removes crosslinks and replication-blocking lesions^{26–29}, RFs encountering an R-loop could target the action of BRCA2, BRCA1 and possibly other Fanconi anaemia proteins to prevent fork collapse and reversal, probably impeding R-loop extension, and could promote re-starting of the replication fork and dissolution of the R-loop (Extended Data Fig. 9b). Thus, TREX-2 could function at co-transcriptional R-loops encountering RFs by regulating the stability and/or recruitment of BRCA2 to the ssDNA substrates generated at the sites of collisions³⁰. Our results reveal a new and unexpected role for tumour suppressors in preventing R-loops, suggesting that R-loops may be a chief cause of replication stress and hence cell death and cancer.

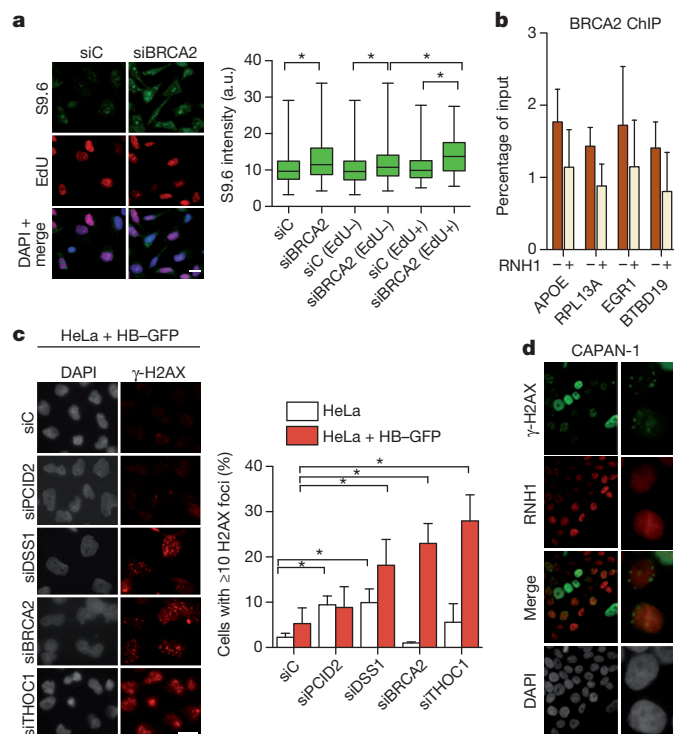


Figure 4 | R-loops trigger genome instability in BRCA2-deficient cells. **a**, Immunofluorescence using the S9.6 antibody in siRNA-treated HeLa cells, with cells in S phase labelled with the nucleotide analogue 5-ethynyl-2'-deoxyuridine (EdU). Quantification of S9.6 nuclear signal. Median, 25th to 75th centile range (boxes) and minimum to maximum centile range (whiskers) are plotted ($n = 3$). $*P \leq 0.05$ (Mann-Whitney test). **b**, ChIP using anti-BRCA2 antibody in HeLa cells ($n = 2$). **c**, γ -H2AX foci in siRNA-transfected HeLa cells expressing HB-GFP ($n = 3$). $*P \leq 0.05$ (two-tailed Student's *t*-tests). Means and s.e.m. are plotted in **b** and **c**. **d**, Immunostaining with anti-RNH1 and anti- γ -H2AX antibodies in BRCA2-deficient CAPAN-1 cancer cells transfected with the RNH1-encoding plasmid ($n = 3$).

METHODS SUMMARY

ON-TARGET SMART pools of standard siRNAs were used from Thermo Scientific. DNA damage foci were analysed by immunofluorescence and quantified with Metamorph (Molecular Probes) image analysis software. Scale bars shown in all images are 20 μ m. DNA combing was performed as described¹¹, with minor modifications. For HB-GFP-retention analysis, HeLa cells were permeabilized before fixation with 0.05% Triton X-100 and were processed for microscopy or fluorescence-activated cell sorting (FACS). For staining with S9.6, fixed cells were either pretreated with 0.5% SDS or pretreated with ice-cold methanol followed by washing with acetone. DRIP was performed as described²². For ChIP with HB-GFP, HeLa cells carrying the *tet*-regulated HB-GFP fusion were used. After 48 h of transfection with siRNA, HB-GFP expression was induced by treatment with 2 μ g ml⁻¹

doxycycline for 24 h. For RNH1-positive samples, cells were transfected with the RNH1-encoding plasmid and processed for ChIP. The number of biological repeats (*n*) performed in each case is indicated in the figure legends.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 October 2013; accepted 8 April 2014.

Published online 1 June 2014.

- Aguilera, A. & García-Muse, T. R loops: from transcription byproducts to threats to genome stability. *Mol. Cell* **46**, 115–124 (2012).
- Huertas, P. & Aguilera, A. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol. Cell* **12**, 711–721 (2003).
- Tous, C. & Aguilera, A. Impairment of transcription elongation by R-loops *in vitro*. *Biochem. Biophys. Res. Commun.* **360**, 428–432 (2007).
- Kaneko, S., Chu, C., Shatkin, A. & Manley, J. Human capping enzyme promotes formation of transcriptional R loops *in vitro*. *Proc. Natl Acad. Sci. USA* **104**, 17620–17625 (2007).
- Skourti-Stathaki, K., Proudfoot, N. & Gromak, N. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell* **42**, 794–805 (2011).
- Mischo, H. *et al.* Yeast Sen1 helicase protects the genome from transcription-associated instability. *Mol. Cell* **41**, 21–32 (2011).
- Tuduri, S. *et al.* Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nature Cell Biol.* **11**, 1315–1324 (2009).
- Wellinger, R., Prado, F. & Aguilera, A. Replication fork progression is impaired by transcription in hyperrecombinant yeast cells lacking a functional THO complex. *Mol. Cell. Biol.* **26**, 3327–3334 (2006).
- Gan, W. *et al.* R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Dev.* **25**, 2041–2056 (2011).
- Helmrich, A., Ballarín, M. & Tora, L. Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Mol. Cell* **44**, 966–977 (2011).
- Domínguez-Sánchez, M., Barroso, S., Gómez-González, B., Luna, R. & Aguilera, A. Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. *PLoS Genet.* **7**, e1002386 (2011).
- Castellano-Pozo, M., García-Muse, T. & Aguilera, A. R-loops cause replication impairment and genome instability during meiosis. *EMBO Rep.* **13**, 923–929 (2012).
- Gómez-González, B. *et al.* Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. *EMBO J.* **30**, 3106–3119 (2011).
- Jani, D. *et al.* Functional and structural characterization of the mammalian TREX-2 complex that links transcription with nuclear messenger RNA export. *Nucleic Acids Res.* **40**, 4562–4573 (2012).
- Cabal, G. *et al.* SAGA interacting factors confine sub-diffusion of transcribed genes to the nuclear envelope. *Nature* **441**, 770–773 (2006).
- Umlauf, D. *et al.* The human TREX-2 complex is stably associated with the nuclear pore basket. *J. Cell Sci.* **126**, 2656–2667 (2013).
- Gallardo, M., Luna, R., Erdjument-Bromage, H., Tempst, P. & Aguilera, A. Nab2p and the Thp1p–Sac3p complex functionally interact at the interface between transcription and mRNA metabolism. *J. Biol. Chem.* **278**, 24225–24232 (2003).
- Ellisdon, A., Dimitrova, L., Hurt, E. & Stewart, M. Structural basis for the assembly and nucleic acid binding of the TREX-2 transcription-export complex. *Nature Struct. Mol. Biol.* **19**, 328–336 (2012).
- Li, J. *et al.* DSS1 is required for the stability of BRCA2. *Oncogene* **25**, 1186–1194 (2006).
- Söderberg, O. *et al.* Direct observation of individual endogenous protein complexes *in situ* by proximity ligation. *Nature Methods* **3**, 995–1000 (2006).
- El Hage, A., French, S., Beyer, A. & Tollervey, D. Loss of topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev.* **24**, 1546–1558 (2010).
- Ginno, P., Lott, P., Christensen, H., Korf, I. & Chédin, F. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol. Cell* **45**, 814–825 (2012).
- Urtishak, K. *et al.* Timeless maintains genomic stability and suppresses sister chromatid exchange during unperturbed DNA replication. *J. Biol. Chem.* **284**, 8777–8785 (2009).
- Gravells, P. *et al.* Reduced FANCD2 influences spontaneous SCE and RAD51 foci formation in uveal melanoma and Fanconi anaemia. *Oncogene* **32**, 5338–5346 (2013).
- Bermejo, R. *et al.* The replication checkpoint protects fork stability by releasing transcribed genes from nuclear pores. *Cell* **146**, 233–246 (2011).
- Lomonosov, M., Anand, S., Sangrithi, M., Davies, R. & Venkitaraman, A. Stabilization of stalled DNA replication forks by the BRCA2 breast cancer susceptibility protein. *Genes Dev.* **17**, 3017–3022 (2003).
- Schlacher, K. *et al.* Double-strand break repair-independent role for BRCA2 in blocking stalled replication fork degradation by MRE11. *Cell* **145**, 529–542 (2011).
- Schlacher, K., Wu, H. & Jasin, M. A distinct replication fork protection pathway connects Fanconi anemia tumor suppressors to RAD51-BRCA1/2. *Cancer Cell* **22**, 106–116 (2012).
- Moldovan, G.-L. & D'Andrea, A. How the Fanconi anemia pathway guards the genome. *Annu. Rev. Genet.* **43**, 223–249 (2009).
- Yang, H. *et al.* BRCA2 function in DNA binding and recombination from a BRCA2–DSS1–ssDNA structure. *Science* **297**, 1837–1848 (2002).

Acknowledgements We thank J.C. Reyes and A.G. Rondón for comments on the manuscript, and D. Haun for style supervision. Research was funded by grants from the Spanish Ministry of Economy and Competitiveness (Consolider CSD2007-00015 and BFU2010-16372), the Junta de Andalucía (CVI4567) and the European Union (FEDER).

Author Contributions V.B., S.B., M.G.R., E.T. and E.H.M. performed the experiments. V.B. and A.A. designed the experiments and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.A. (aguilo@us.es).

METHODS

Antibodies. Anti- γ -H2AX (clone JBW301; Upstate), anti-53BP1 (NB100-304; Abnityc Biopharma), anti-RAD51 (ab213; Abcam), anti-PCID2 (SC-84568; Santa Cruz), anti-TOP1 (ab3825; Abcam), anti-BRCA2 (OP95; Calbiochem), antiDSS1 (136391-AP; Proteintech), anti-RNH1 (156061-AP; Proteintech), anti-BRCA1 (ab16780; Abcam), anti-RPA32 (ab-2175; Abcam), anti-RAD51 (ab-213; Abcam) and anti-GFP (11814460001; Roche) antibodies were used. Anti-GANP was a gift from V. Wickramasinghe.

siRNA depletion and mRNA quantification. ON-TARGET SMARTpool of siRNA from Thermo Scientific were used for all depletions. Lipofectamine 2000 (Invitrogen) was used for transfections in accordance with the manufacturer's instructions, and cells were used 72 h after transfection. Complementary DNA from cytoplasmic RNA (1 mg) was generated by reverse transcription using Super-Script first-strand synthesis for PCR with reverse transcription (Invitrogen) and random primers. Real-time quantitative PCR (qPCR) was performed with SYBR qPCR Mix (Applied Biosystems) and analysed on an ABI Prism 7000 (Applied Biosystems).

RNA primers for real-time qPCR. The following primers were used for real-time qPCR: APOE, 5'-CCGGTGAAGCGCAGTCGG-3' (forward) and 5'-CCCAA GCCGACCCCGAGTA-3' (reverse); RPL13A, 5'-GCTTCCAGCACAGGACA GGTAT-3' (forward) and 5'-CACCCACTACCCGAGTTCAAG-3' (reverse); EGRI, 5'-TTCGGATTCCCGCAGTGT-3' (forward) and 5'-TCACTTTCCCCCTTT ATCCA-3' (reverse); BTBD19, 5'-CCCCAAAGGGTGGTGA-3' (forward) and 5'-TTCACATTACCCAGACCAGACTGT-3' (reverse); SNRPN, 5'-TGCC AGGAAGCCAAATGAGT-3' (forward) and 5'-TCCCTCTTGGAACATCCA-3' (reverse); LIG_P1, 5'-GGCTGCGGCAGTTGTGA-3' (forward) and 5'-CAATG CAGCTTTGAGGAAACC-3' (reverse); LIG_P2, 5'-TCTGAGGGTGGAAACC ATACAA-3' (forward) and 5'-CAAATCTGCCTTTTGAGAACCA-3' (reverse); LIG_P3, 5'-CTGACCTGTTCTGTTTGGATT-3' (forward) and 5'-TTCCCA ATACCAGCCCTTT-3' (reverse); UTRN_P1, 5'-GGCAAGATGGCCAAGTAT GGAG-3' (forward) and 5'-GCTTTCTTGAGCTTCCCTTACCTACCAG-3' (reverse); UTRN_P2, 5'-TGGATGCTCTCATCGGGAGAA-3' (forward) and 5'-GCAC ACAGGGCAAACACAGGTA-3' (reverse); UTRN_P3, 5'-GGCTACTATGCTT CAACATCGACTGG-3' (forward) and 5'-GTGGTAAGGCTGCGCTTCTCT-3' (reverse); ACTB_P1, 5'-CGGCTGGGTAGGTTTGTAG-3' (forward) and 5'-GGCTTGAGAGGTAGAGTGTG-3' (reverse); ACTB_P2, 5'-CGGGGTCTTTG TCTGAGC-3' (forward) and 5'-CAGTTAGCGCCCAAAGGAC-3' (reverse); ACTB_P3, 5'-TAACACTGGCTCGTGTGACAA-3' (forward) and 5'-AAGTGCAAAG AACACGGCTAA-3' (reverse); ACTB_P4, 5'-CTAAGTCTGCCCTATTTC C-3' (forward) and 5'-GATGTGACAGCTCCCCAC-3' (reverse); BRCA2, 5'-A GGACTTGCCCTTTTCGTTCTA-3' (forward) and 5'-TGCAGCAATTAACAT ATGAGG-3' (reverse); PCID2, 5'-CAGA AGCTGGTGGTGCAGCA-3' (forward) and 5'-GGCTCCGTGTACTTTCAACACA-3' (reverse); DSS1, 5'-GCAGCCGG TAGACTTAGGTCTGT-3' and 5'-TCTTCGGCAGGGAAGTCTTC-3' (reverse).

Western blot analysis. Western blots were performed in the standard way for most proteins; that is, using 4–20% gradient gel and Tris–glycine transfer buffer. For DSS1, 20% SDS–PAGE gel was used to resolve proteins, and 25 mM KH_2HPO_4 buffer was used for transfer. Fixation was performed for 45 min with 0.2% (v/v) glutaraldehyde (ref. 19). For GANP and BRCA2, 5% SDS–PAGE gels were used.

Immunofluorescence microscopy. For analysis of DNA damage foci, cells were fixed with 2% formaldehyde, washed and treated with ice-cold 70% ethanol, blocked with 2% BSA in PBS and stained with primary antibody (anti- γ -H2AX (1:400 dilution), anti-53BP1 (1:400) or anti-RAD51 (1:200)), and washed and stained with Alexa Fluor-conjugated secondary antibodies (1:1,000) (Life Technologies). Images were captured at $\times 63$ magnification with a Leica wide-field microscope. Metamorph (Molecular Probes) image analysis software was used to quantify foci. For PCID2 subcellular localization, cells were pre-permeabilized for 5 min with the indicated concentration of Triton X-100, and fixed with 2% formaldehyde. When stated, cells expressing HB–GFP were pre-permeabilized for 5 min with 0.1% Triton X-100 before fixation with formaldehyde; they were treated with ice-cold ethanol after fixation. For indirect immunofluorescence with monoclonal S9.6 antibody, cells were specifically treated with 0.5% SDS after fixation, or fixed and permeabilized with ice-cold methanol for 10 min and acetone for 1 min on ice. For fluorescence quantification analysis of the S9.6 signal in S-phase cells, cells were incubated for 20 min with 5-ethynyl-2'-deoxyuridine. Click-iT 5-ethynyl-2'-deoxyuridine Alexa Fluor 555 Imaging Kit was used to detect cells in S phase. Cells were blocked with 3% BSA, 0.1% Tween 20 in $4 \times \text{SSC}$ and incubated with primary anti-S9.6 antibody (1:100 dilution) and secondary chicken anti-mouse Alexa Fluor 488 (1:500; Invitrogen) antibodies. The scale bar shown on the bottom right corner of each image represents 20 μm .

Replication analysis by DNA combing. DNA combing was performed as described¹¹, except that both iododeoxyuridine and chlorodeoxyuridine labels were added for 10 min each.

Proximity ligation assay. The proximity ligation assay was performed with reagents from Olink Biosciences in accordance with the manufacturer's instructions. For negative controls, everything was performed identically, except that only one of the primary antibodies was added.

Construction of the HB–GFP fusion. pcDNA3-RNaseH1 (ref. 31) was used to get the HB domain of RNH1 that was cloned using primers RNH1_HBF (5'-ACTCAG TATCTGGGATGTTCTATGCCGTGAGG-3') and RNH1_HBR (5'-ATTGAG TCGACGCTTGCTGATTCTCTGAC-3') into pEGFP-C1 vector (Clontech). To create stable HeLa cell lines used for ChIP experiments, HB-domain-tagged EGFP (HB–GFP) from pEGFP-C1 was cloned into pT-Rex-DEST30 Gateway vector, downstream of the TetO2 operator. pT-Rex-DEST30-HB–GFP was transfected to a HeLa stable cell line carrying pcDNA6TR (TetR expression vector).

Size-exclusion chromatography. HeLa cells were lysed with 1% Triton X-100 and 0.5% Nonidet P40 in PBS with $1 \times$ protease inhibitor cocktail. The lysate was centrifuged after a brief sonication and loaded on a pre-equilibrated Superose 6 column (17-5172-01; Gelifesciences). Fractions obtained were precipitated with trichloroacetic acid, washed with acetone and analysed by western blotting.

HB–GFP-retention FACS assay. Cells were treated with trypsin and resuspended in ice-cold PBS containing 1% FBS, then permeabilized for 4 min with 0.05% Triton X-100 in presence of $1 \times$ protease inhibitor cocktail (Roche). Three volumes of 100% ice-cold ethanol was added and kept in ice for 1 h, treated in a standard manner with propidium iodide and RNase, and analysed by FACS.

Chromatin immunoprecipitation (ChIP). HB–GFP-expressing HeLa cells were transfected with siRNA, and HB–GFP expression was induced after 48 h. For RNH1-positive samples the RNH1 encoding plasmid was transfected after 48 h of siRNA transfection. After 72 h of siRNA transfection, cells were crosslinked and processed for ChIP with the use of standard procedures with minor modifications. In brief, cells were crosslinked for 10 min with 1% formaldehyde, washed with dilution buffer (50 mM Tris-HCl pH 7.5, 250 mM NaCl, 1 mM EDTA) and resuspended in 1 ml of lysis buffer (50 mM Tris pH 7.5, 250 mM NaCl, 1 mM EDTA, 1% Triton X-100, 1% Nonidet P40, 1% SDS, $1 \times$ protease inhibitor cocktail) and sonicated on the maximum intensity setting, with ten pulses of 30 s on and 1 min off in Bioruptor (Diagenode), to obtain approx. 800-kilobase fragments. The lysate was centrifuged, and 20 and 200 μl of supernatant were used for input and immunoprecipitation, respectively. GFP–Trap-M (Chromotex) was used for HB–GFP immunoprecipitation. Cells were washed with wash buffer 1 (50 mM Tris-HCl pH 7.5, 250 mM NaCl, 1 mM EDTA, 0.5% Triton X-100, 0.5% Nonidet P40, 0.01% SDS, $1 \times$ protease inhibitor cocktail) and wash buffer 2 (50 mM Tris-HCl pH 7.5, 250 mM LiCl, 1 mM EDTA, 0.5% sodium deoxycholate, 0.5% Nonidet P40, 0.01% SDS, $1 \times$ protease inhibitor cocktail). Input and immunoprecipitate samples were then un-crosslinked and treated with protease K before DNA isolation with the QIAamp DNA Mini Kit (Qiagen), then subjected to qPCR with the primers listed above. The dilution factor was adjusted and the percentage of the input signal was calculated. For BRCA2 ChIP, Dynabeads Protein G (Life Technologies), was incubated with anti-BRCA2 (OP-95), washed and used for immunoprecipitation.

DRIP in HeLa cells. DRIP analysis was performed as described²². In brief, 5×10^6 HeLa cells were collected, washed with PBS, resuspended in 1.6 ml of Tris–EDTA (TE) buffer and treated overnight with 41.5 μl of 20% SDS and 5 μl of proteinase K (Roche). DNA was extracted gently with phenol–chloroform. Precipitated DNA was spooled on a glass rod, washed with 70% ethanol, resuspended gently in TE and digested overnight with 50 U of HindIII, EcoRI, BsrGI, XbaI and SspI, 2 mM spermidine and BSA. As negative control, half of the DNA was treated overnight with 3 μl of RNase H (M0297; New England BioLabs). Digested DNA (5 μg) was bound overnight to 10 μl of S9.6 antibody (1 mg ml^{-1}) in 500 μl of binding buffer (10 mM NaPO_4 , 140 mM NaCl, 0.05% Triton X-100) at 4 °C. DNA–antibody complexes were immunoprecipitated for 2 h with Dynabeads Protein A (Invitrogen) at 4 °C and washed three times with binding buffer. DNA was eluted with 50 mM Tris-HCl pH 8.0, 10 mM EDTA, 0.5% SDS, then treated for 45 min with 7 μl of proteinase K at 55 °C and cleaned with phenol–chloroform. qPCR was performed at the indicated regions. The signal intensity plotted is the relative abundance of DNA–RNA hybrid immunoprecipitated in each region, normalized to input values and to the signal at the SNRPN negative control region (see list of primers above).

RNaseH1-dependent release of HB–GFP from chromatin of HeLa cells. Transfected cells expressing HB–GFP were washed with 0.2% Triton X-100, treated with 2% formaldehyde, lysed for 15 min in lysis buffer (1% Triton X-100, 0.2% SDS, 50 mM Tris-HCl pH 8, 50 mM NaCl, 5 mM MgCl_2 , 1 mM dithiothreitol, 20 μg of BSA, 5% glycerol) and sonicated and split into two aliquots. RNaseH1 (NEB M0297) was added to one half, then incubated at 22 °C for 30 min and subjected to western blot analysis.

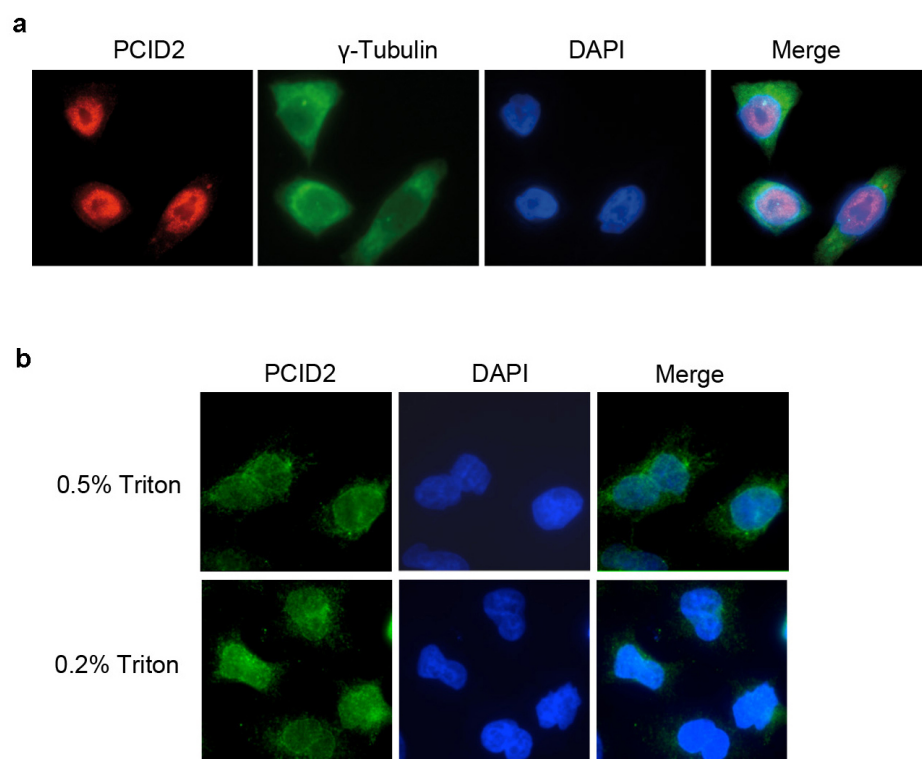
Metaphase spread and sister chromatid exchange analysis. The sister chromatid exchange assay was performed as described³², with minor modifications. In brief, cells were transfected with the indicated plasmid and incubated for 42 h with 10 μM bromodeoxyuridine followed by treatment for 3 h with 0.1 $\mu\text{g ml}^{-1}$ Colcemid.

Cells were harvested at 48 h after plasmid transfection. Metaphase spreads were prepared and the slides were incubated for 20 min with $20 \mu\text{g ml}^{-1}$ Hoechst solution, then exposed for 1 h to ultraviolet A in and incubated for 20 min in $2 \times \text{SSC}$ before Giemsa staining was performed. Metaphases were scored with a $100\times$ objective, and at least 25 images were taken randomly from each condition.

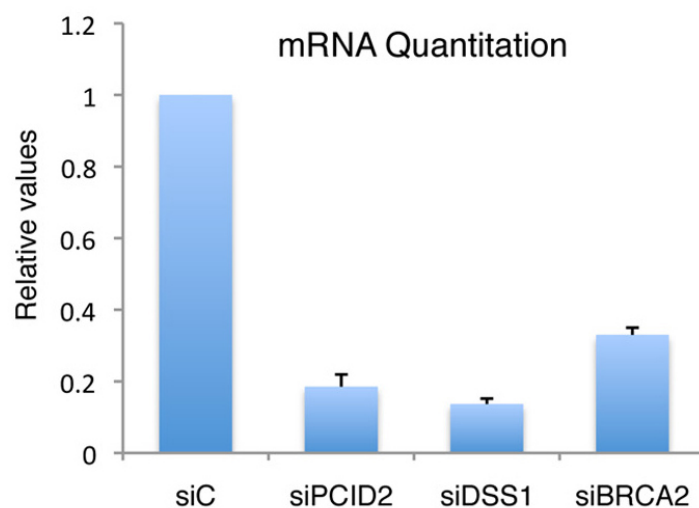
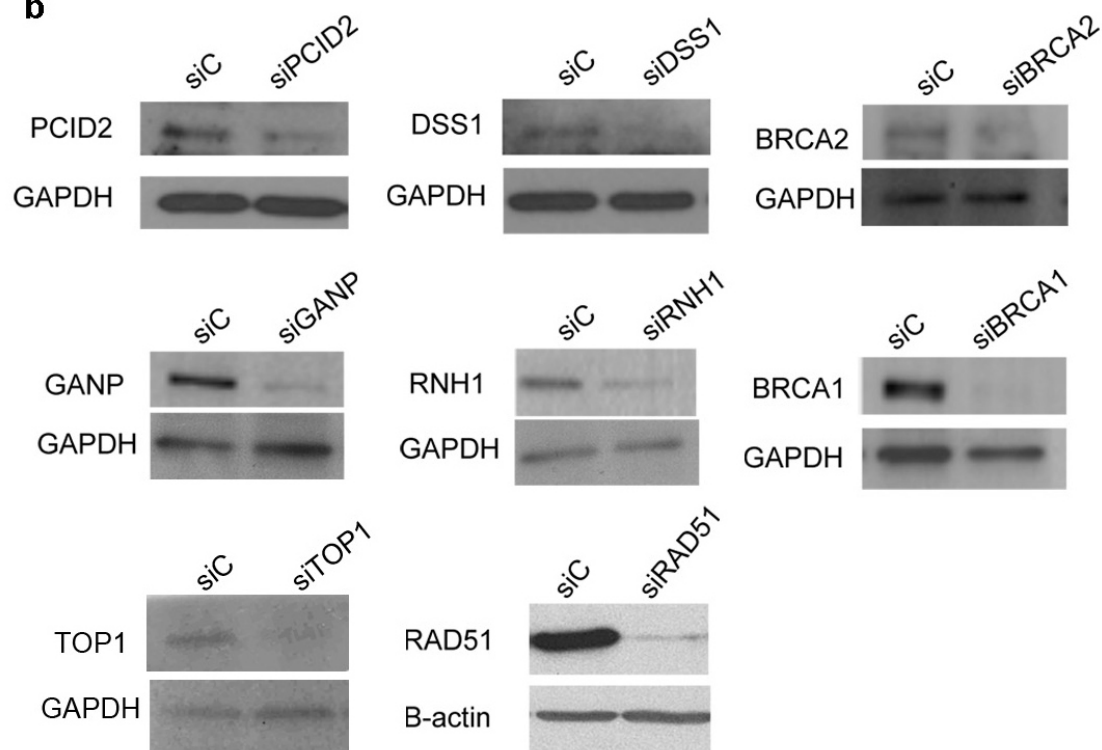
Adhesion-independent cell proliferation. RPE-1 cells were used for adhesion-independent cell growth; 0.5% DNA-grade agarose/DMEM F-12 medium was added to 96-well plates to prevent cell attachment to and cell-monolayer formation on the plastic base. The plates were stored at 4°C . RPE cells were transfected with siRNA and/or plasmid as indicated. After 72 h, on day 4, cells were dissociated with trypsin-EDTA (Gibco) and resuspended in medium. Cells were counted, centrifuged, and resuspended in complete DMEM F-12 medium with 0.3% agarose, prepared at

42°C , and dispensed into a 96-well plate prewarmed in a 37°C incubator. Cells were incubated at 37°C and quantified on day 9 using WST-1 reagent (Roche) in accordance with the manufacturer's protocol. Absorbance was read in a Varioskan Flash Multimode Reader.

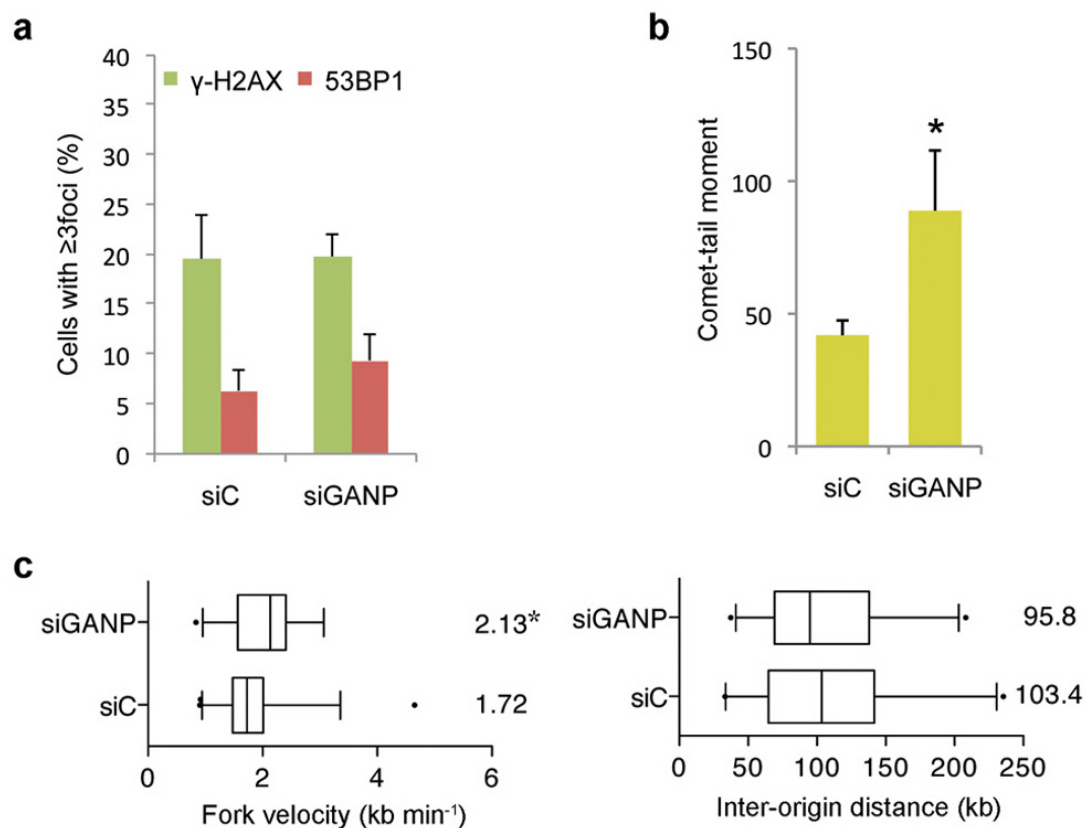
31. ten Asbroek, A., van Groenigen, M., Nooij, M. & Baas, F. The involvement of human ribonucleases H1 and H2 in the variation of response of cells to antisense phosphorothioate oligonucleotides. *Eur. J. Biochem.* **269**, 583–592 (2002).
32. Bayani, J. & Squire, J. A. Sister chromatid exchange. *Curr. Protoc. Cell Biol.* **25**, 22.7.1–22.7.4 (2005).
33. Smith, A., Friedman, D., Yu, H., Carnahan, R. & Reynolds, A. ReCLIP (reversible cross-link immuno-precipitation): an efficient method for interrogation of labile protein complexes. *PLoS ONE* **6**, e16206 (2011).



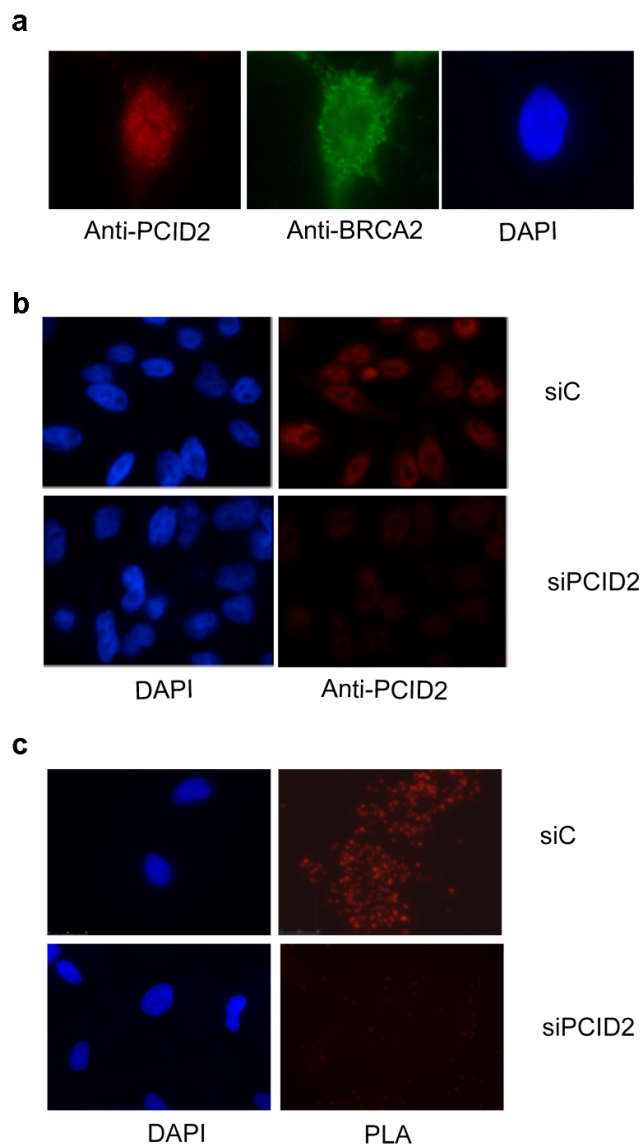
Extended Data Figure 1 | Subcellular localization of PCID2. Immunofluorescence of endogenous PCID2 in HeLa cells. **a**, Without permeabilization. **b**, With pre- permeabilization (see Methods).

a**b**

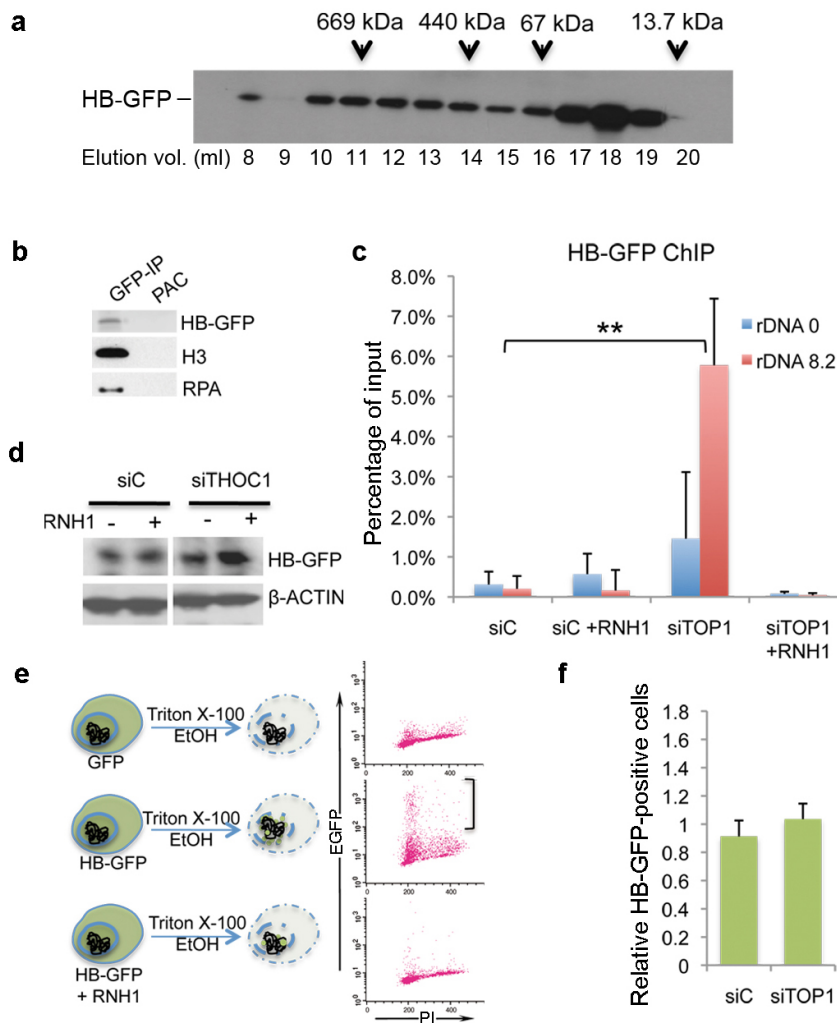
Extended Data Figure 2 | Validation of siRNAs. **a**, Relative mRNA quantification. Means and s.e.m. are plotted. **b**, Western blot analysis of siRNA-treated HeLa cells.



Extended Data Figure 3 | Effect of GANP depletion in genomic instability. **a**, γ -H2AX and 53BP1 foci. **b**, Single-cell electrophoresis. **c**, DNA-combing analysis in GANP-depleted cells. Details as in Fig. 1.

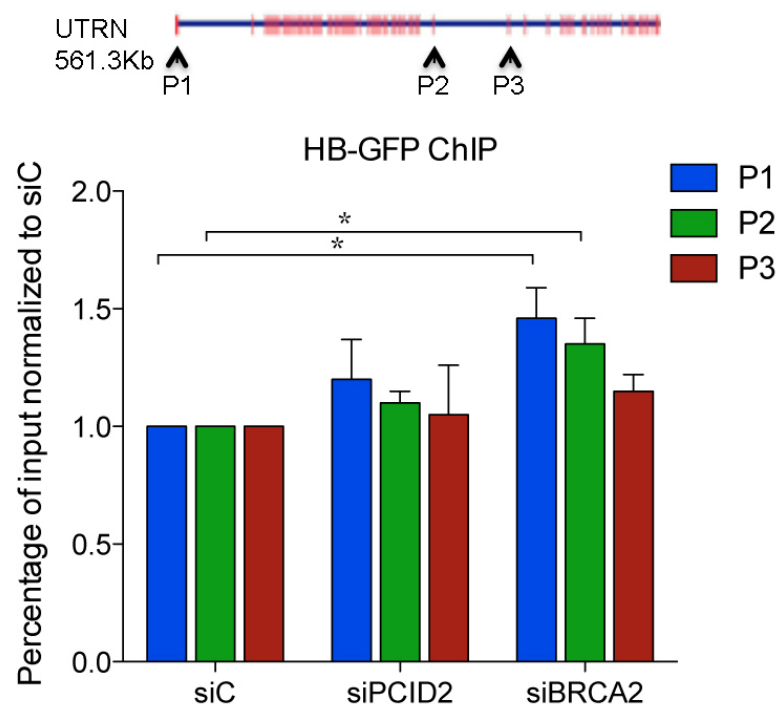
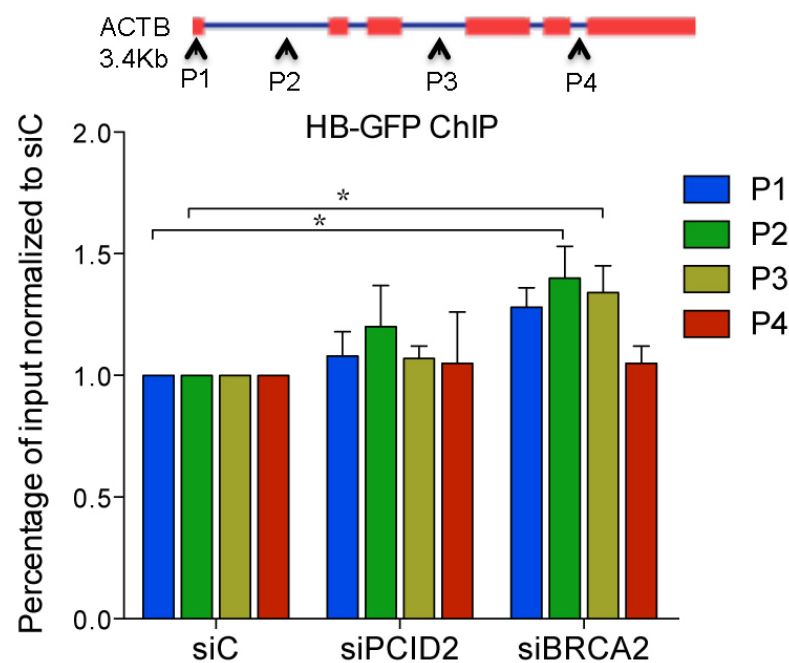


Extended Data Figure 4 | Proximity ligation assay in PCID2-depleted cells.
a, Staining pattern observed with both anti-BRCA2 and anti-PCID2 antibodies in conditions used for proximity ligation assay in Fig. 1d. **b**, Anti-PCID2 immunofluorescence analysis. **c**, Proximity ligation assay in control and PCID2 siRNA-treated cells (see Fig. 1d).



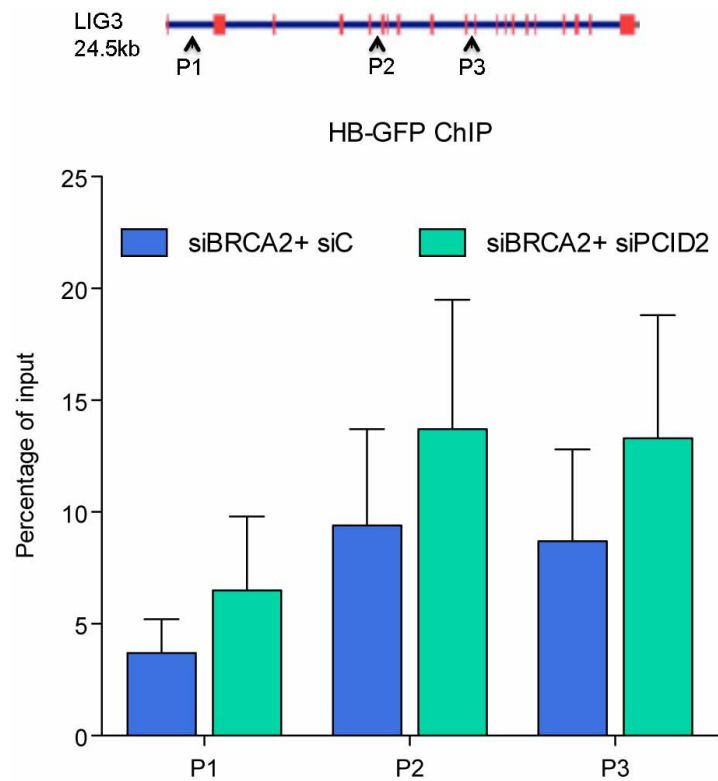
Extended Data Figure 5 | HB-GFP interacts with chromatin and chromatin-associated proteins by means of DNA-RNA hybrids. **a**, HB-GFP-expressing HEK293 lysate fractionated on Superose 6 size-exclusion columns (17-5172-01; Gelifesciences) and analysed by western blotting. **b**, HB-GFP co-immunoprecipitated proteins by using the ReCLIP method³³. **c**, HB-GFP ChIP in the ribosomal DNA region in TOP1-depleted HeLa cells

($n = 3$). Means and s.e.m. are plotted. **d**, RNH1-dependent release of HB-GFP from chromatin of HeLa cells. **e**, Scheme and representative plot of FACS assays used to quantify DNA-RNA hybrids. PI, propidium iodide (see Fig. 2f). **f**, FACS assay to quantify DNA-RNA hybrids in TOP1-depleted cells. Means and s.e.m. are plotted ($n = 3$).

a**b**

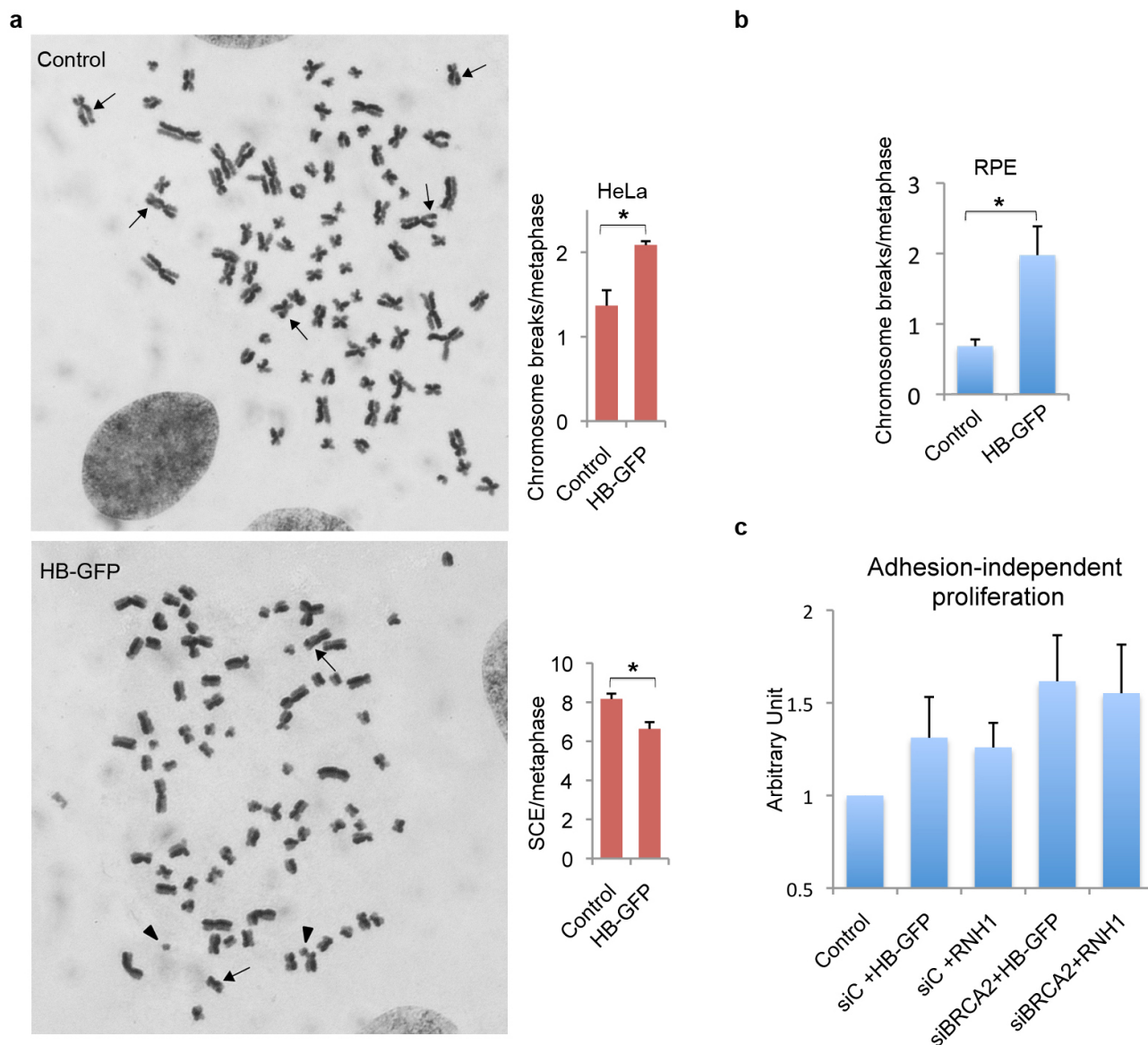
Extended Data Figure 6 | HB-GFP ChIP of the *UTRN* and *ACTB* gene regions. **a, b,** Normalized siRNA control-relative values for the immunoprecipitated DNA-RNA hybrids plotted relative to the siRNA control

in the *UTRN* (**a**) and *ACTB* (**b**) genes. Means and s.e.m. are shown ($n = 3$). Amplicon positions used are indicated in the gene diagram above each graph.



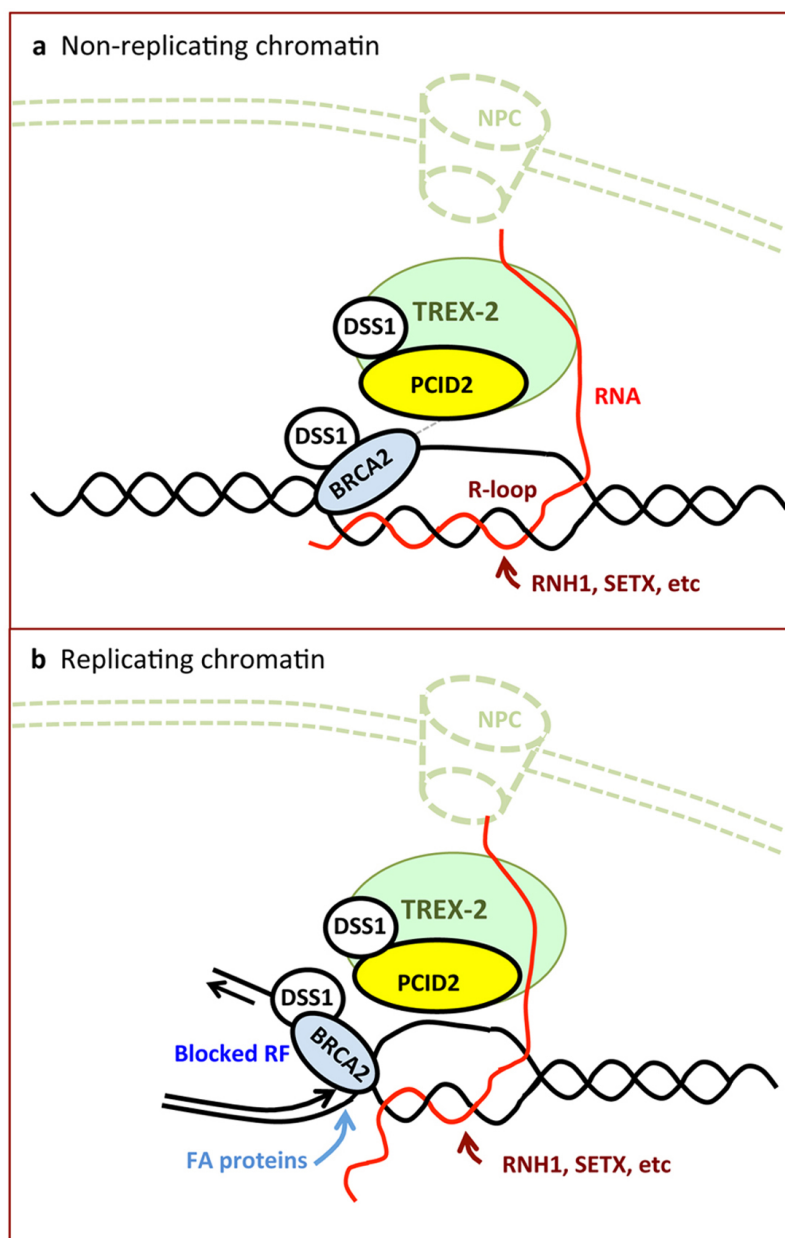
Extended Data Figure 7 | Accumulation of DNA–RNA hybrids in cells depleted of both PCID2 and BRCA2. HeLa cells were treated with siBRCA2

in combination with siRNA control or siPCID2, and processed. Means and s.e.m. are plotted ($n = 3$). Details as in Fig. 3b.



Extended Data Figure 8 | Chromosomal aberrations in cells expressing HB-GFP. **a**, Metaphase spreads of HeLa cells expressing GFP (control) or HB-GFP. Fragmentation and sister chromatid exchange events are indicated by arrowheads and arrows, respectively. **b**, Quantification of chromosome

breaks in RPE cells expressing HB-GFP. **c**, Adhesion-independent proliferation assay. Cell proliferation relative to control siRNA-treated RPE cells is shown. Means and s.e.m. are plotted ($n = 3$). $*P \leq 0.05$ (two-tailed Student's t -test).



Extended Data Figure 9 | Model to explain the role of BRCA2 preventing R-loops as a source of genome instability. **a**, RNA–DNA hybrids may form both in the interior and at the periphery of the nucleus. mRNP biogenesis factors such as the TREX-2 complex may help recruit or stabilize BRCA2 near transcribed regions, whether or not these are in proximity to the nuclear pore complex. BRCA2 and other related proteins could bind to the branched structure generated by the ssDNA displaced in the R-loop, facilitating the

action of enzymes that remove R-loops, such as specific RNases or DNA–RNA helicases. This could occur in non-replicating chromatin. **b**, In replicating chromatin, BRCA2 and, presumably, other Fanconi anaemia proteins may act directly at putatively stalled RFs in front of an R-loop to impede the collapse or reversal of the replication fork, probably impeding R-loop extension. Subsequently, R-loop removal could be promoted by the passage of the replication fork.

The structural basis of transfer RNA mimicry and conformational plasticity by a viral RNA

Timothy M. Colussi^{1,2†}, David A. Costantino^{1,2}, John A. Hammond^{1†}, Grant M. Ruehle¹, Jay C. Nix³ & Jeffrey S. Kieft^{1,2}

RNA is arguably the most functionally diverse biological macromolecule. In some cases a single discrete RNA sequence performs multiple roles, and this can be conferred by a complex three-dimensional structure. Such multifunctionality can also be driven or enhanced by the ability of a given RNA to assume different conformational (and therefore functional) states¹. Despite its biological importance, a detailed structural understanding of the paradigm of RNA structure-driven multifunctionality is lacking. To address this gap it is useful to study examples from single-stranded positive-sense RNA viruses, a prototype being the tRNA-like structure (TLS) found at the 3' end of the turnip yellow mosaic virus (TYMV). This TLS not only acts like a tRNA to drive aminoacylation of the viral genomic (g)RNA^{2–4}, but also interacts with other structures in the 3' untranslated region of the gRNA⁵, contains the promoter for negative-strand synthesis, and influences several infection-critical processes⁶. TLS RNA can provide a glimpse into the structural basis of RNA multifunctionality and plasticity, but for decades its high-resolution structure has remained elusive. Here we present the crystal structure of the complete TYMV TLS to 2.0 Å resolution. Globally, the RNA adopts a shape that mimics tRNA, but it uses a very different set of intramolecular interactions to achieve this shape. These interactions also

allow the TLS to readily switch conformations. In addition, the TLS structure is 'two faced': one face closely mimics tRNA and drives aminoacylation, the other face diverges from tRNA and enables additional functionality. The TLS is thus structured to perform several functions and interact with diverse binding partners, and we demonstrate its ability to specifically bind to ribosomes.

The TYMV TLS RNA (hereafter termed 'the TLS') is a tRNA mimic, a subject of broad biological and evolutionary importance⁷, as highlighted by the fact that some tRNA mimics are linked to disease^{8–10}. Like tRNA, the aminoacylated TLS binds to eukaryotic elongation factor 1A (eEF1A) and is a substrate for tRNA-modifying enzymes⁶. These activities and other data suggest a tRNA-like structure^{11–16}. However, the topology of the TLS differs from tRNA, mandated by its location on the 3' end of the gRNA (Fig. 1a, b and Extended Data Fig. 1). In addition to affecting many viral processes^{17–19}, the TLS may regulate the activities of ribosomes and replicases on the gRNA^{6,20}. This function could be conferred by its ability to readily transition between folded and unfolded states. Simple tRNA mimicry is insufficient to explain these phenomena; although tRNAs flex while transiting through the ribosome they do not unfold and refold. To explore the paradigms of tRNA mimicry and RNA structural and functional plasticity, we solved the structure

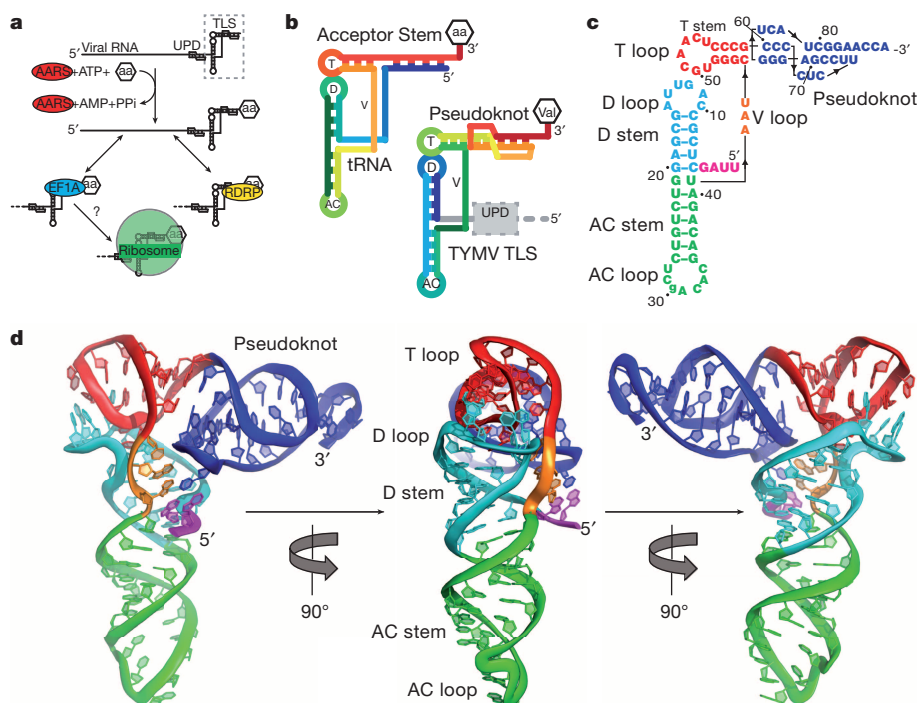


Figure 1 | Function and structure of the TYMV TLS. **a**, The TLS (dashed box) at the 3' end of the gRNA, with the UPD upstream. AARS (red) valylates the TLS, which can interact with the RDRP (yellow) or eEF1A (blue). Ribosome binding (green) was suspected but untested. aa, amino acid. **b**, Topology of tRNA and the TLS in rainbow colours. 5' Ends are blue and 3' ends are red. Attached amino acid is labelled 'aa' or Val (for valine). AC, anticodon loop; D, D loop; T, T loop; V, variable loop. **c**, Secondary structure of the crystallized RNA. Lowercase letter indicates the single mutation. Numbering is from the 5' end of the crystallized sequence. **d**, Three views of the structure, coloured to match scheme used in **c**. The conformation of the 3' CCA and the AC loop differ from tRNA, probably owing to crystal packing (Extended Data Fig. 8).

¹Department of Biochemistry and Molecular Genetics, University of Colorado Denver School of Medicine, Aurora, Colorado 80045, USA. ²Howard Hughes Medical Institute, University of Colorado Denver School of Medicine, Aurora, Colorado 80045, USA. ³Molecular Biology Consortium, Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [†]Present addresses: Department of Chemistry and Chemical Biology, Northeastern University, Boston, Massachusetts 02115, USA (T.M.C.); Department of Integrative Structural and Computational Biology, Scripps Research Institute, La Jolla, California 92037, USA (J.A.H.).

of the TYMV TLS RNA by X-ray crystallography to 2.0 Å resolution (Fig. 1c and Extended Data Fig. 2), comparable to the highest-resolution structure of free tRNA, which we used here for comparison (1.93 Å)²¹.

The TLS assumes the classic L-shaped tRNA conformation (Fig. 1d), but achieves this in a way that diverges from tRNA and from previous predictions^{13,22}. The topology (Fig. 1b and Extended Data Fig. 3) and the intramolecular interactions that form the structure are different from those in tRNA (Fig. 2a). Although the TLS pseudoknot (the first recognized RNA pseudoknot¹⁴) is in the position of the acceptor stem of the tRNA, and elements analogous to the D loop, T loop and V loop are positioned as in tRNA, their interactions are not tRNA-like. In the elbow region of tRNA, the V loop interacts with the D stem, stabilizing the L-shaped tRNA structure (Fig. 2b). In contrast, the V-loop bases in the TLS point away from the D stem to interact with the 5' end and pseudoknot of the TLS (Fig. 2b). G4 adopts a *syn* conformation (Extended Data Fig. 4), forming a long-range base pair with C76 in a loop of the pseudoknot. The unexpected G4–C76 base pair is stabilized by stacking of A3 and the V-loop base A42 on either side. V-loop bases A42–U44

continuously stack to reach from the G4–C76 pair to A15 of the D loop, structurally linking the pair to the elbow region. These interactions explain the observation that removing the 5'-UUAG sequence from the 5' end of the TLS (including G4) destabilizes the L-shape conformation and elbow structure (Extended Data Fig. 5)^{5,11}. Although historically not recognized as part of the minimal TLS, these 5' nucleotides form a 'linchpin' interaction that stabilizes the global structure and this suggests why their presence increases aminoacylation efficiency (Fig. 2a and Extended Data Fig. 6).

The structural features of the elbow region require that the TLS D loop be in a different conformation from that of the tRNA. In the TLS, the D loop contains a tight bend that allows A15 to reach across the helix to stack on U44 in the V loop (Fig. 2c), while U13 and U14 lie against the end of the T loop. No analogous bases or interactions are found in tRNA. Despite this, the T loops of the TLS and tRNA are structurally identical (Extended Data Fig. 6). D-loop bases G12 and A11 dock into the T loop almost identically to the analogous bases of tRNA, although A11 is in a *syn* conformation.

The stabilizing intramolecular interactions of the TLS show how it can adopt different folded states, potentially to organize infection-important activities, achieving structural and functional plasticity. Disruption of the linchpin would lead to a loss of the L-shape fold and a propagated loss of interactions extending from the V loop to the D/T-loop interface. This effect is observed when the base pair and adjacent nucleotide that stack on and stabilize this pair are eliminated by truncating the TLS from the 5' end^{5,11} (Extended Data Fig. 5). This disruption could be induced by loading of the virally encoded RNA-dependent RNA polymerase (RDRP) at the 3' end^{18,23,24}. The subsequent destabilization would create a favourable template for the RDRP and effectively remove competition between the RDRP and the proteins that require the stable fold (for example, aminoacyl tRNA synthetase (AARS)).

The TLS structure has two distinct 'faces'. The tRNA-deviating features are on one side of the structure, where the upstream pseudoknot domain (UPD) and the gRNA connect to the TLS (Fig. 1b and Extended Data Fig. 7). The structure reveals that the UPD is positioned to interact with the 'divergent face' of the TLS. The opposing side of the TLS, the 'tRNA-like face', interacts with the valyl-AARS when the TLS structure is modelled into a tRNA^{Val}•AARS complex structure²⁵ (Fig. 3a, b). The TLS structure is accommodated by the AARS, including the acceptor stem pseudoknot, which has a different structure to that shown by NMR (Extended Data Fig. 6). Like tRNA, the TLS has high crystallographic *B* factors in its anticodon (AC) loop and 3' CCA, suggesting that these can readily undergo structural changes (Fig. 3c, d and Extended Data Fig. 8). In the case of the AC loop, this is important to dock the valine-specifying identity elements in the AC loop onto the protein²⁶. Modelling of the TLS structure onto an elongation factor structure also reveals an interface similar to that formed with tRNA and no obvious steric clash (Extended Data Fig. 6). Because the divergent face does not contact the AARS or eEF1A, the 5' end of the TLS is not occluded by interaction with either protein. Thus, the UPD and viral genome do not interfere with binding (Extended Data Fig. 9), and the precise mimicry of the tRNA-like face explains how the TLS can achieve tRNA-like valylation efficiencies and eEF1A binding affinities²⁷.

The interactions of the TLS with AARS and eEF1A suggest that it could bind to the ribosome, as previously suggested^{6,28}. Ribosome binding would require accommodating the entire TLS structure between the subunits, including elements that deviate from tRNA within the TYMV 3' untranslated region (UTR). We measured binding of TLS-containing RNAs to *Thermus thermophilus* 70S ribosomes, a valid model for tRNA binding assays given the interchangeability of eukaryotic and bacterial tRNAs²⁹. *In vitro* transcribed *Arabidopsis thaliana* tRNA^{Val} bound to the 70S (dissociation constant (K_d) = 0.27 ± 0.05 nM) whereas a 75-nucleotide-long negative control RNA (from bacteriophage *phi29* pRNA) did not ($K_d > 1,000$ nM) (Fig. 4a and Extended Data Fig. 1). Mutation of the tRNA^{Val} D loop to disrupt the global tRNA fold resulted in a 28-fold loss of affinity ($K_d > 7.6 \pm 0.8$ nM) (Fig. 4b), consistent with binding

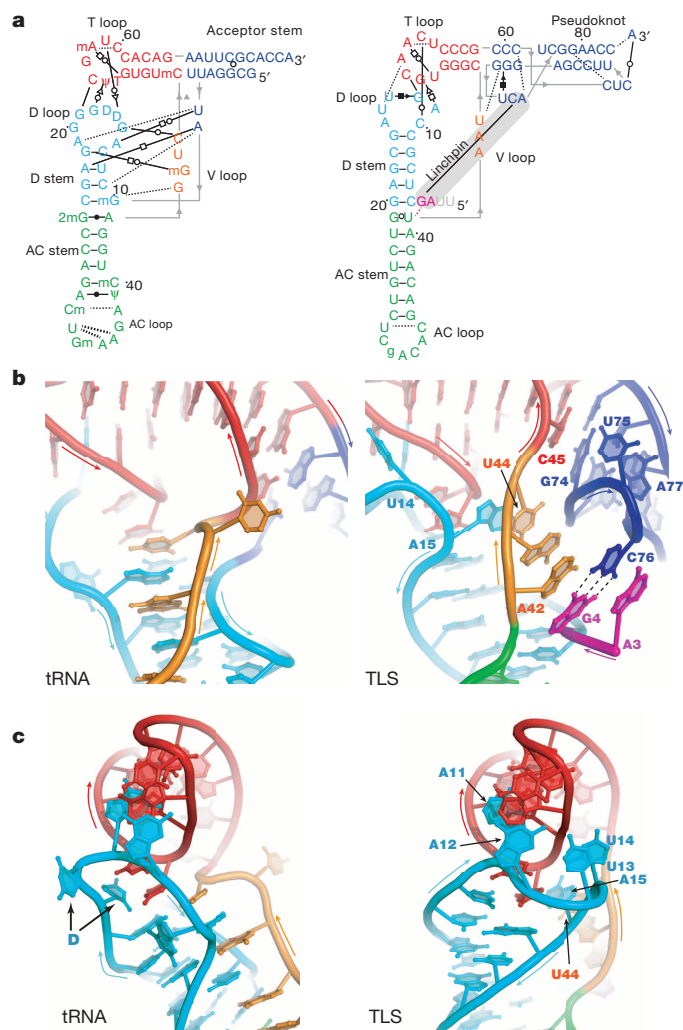


Figure 2 | Structural differences between tRNA and the TLS. **a**, Secondary structures showing interactions that stabilize the folds of tRNA (left) and the TLS (right). Non-canonical base pairs are indicated with Leontis–Westhof symbols³⁰, single hydrogen bonds with dashed lines. Lines with embedded arrows indicate chain connectivity. Grey nucleotides were not visible in the electron density. Grey bar indicates the long-range linchpin interaction. **b**, Intramolecular interactions of the V loop (orange) in tRNA (left) and the TLS (right). Dashed lines indicate the C76–G4 base pair. **c**, Conformation and interaction of the D loop (cyan) with the T loop (red) of tRNA (left) and the TLS (right).

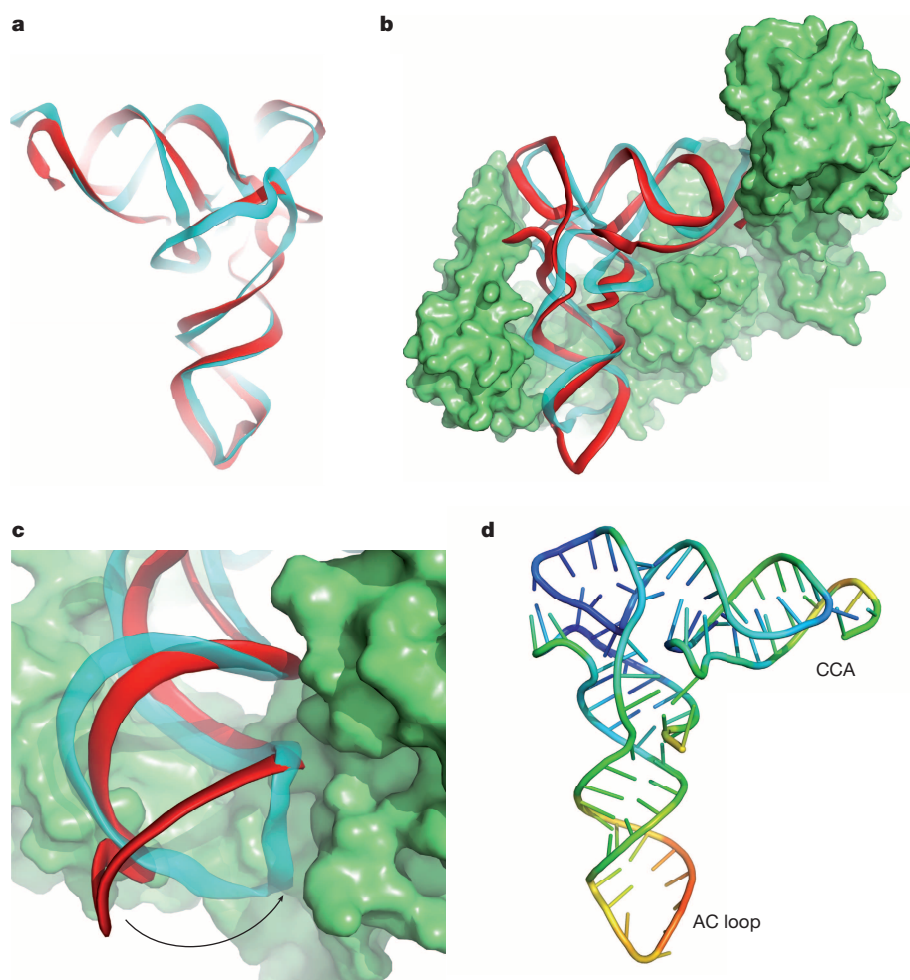


Figure 3 | tRNA mimicry and AARS binding. **a**, Backbone traces of superimposed tRNA (cyan) and TLS (red). The tRNA-like face is shown. **b**, Superposition of the TLS onto tRNA^{Val} bound to valyl-AARS (Protein Data Bank accession 1GAX)²⁵. **c**, The AC loop of the TLS (red) must swing into position to match that of tRNA (cyan). **d**, TLS structure coloured by relative crystallographic *B* factor (high, red; low, blue).

being dependent on the global conformation of the tRNA. A TLS RNA containing the 5'-UUAG sequence bound with tRNA-like affinity ($K_d = 0.31 \pm 0.07$ nM), and mutation of this RNA's D loop decreased binding

ninefold ($K_d > 2.7 \pm 0.2$ nM) (Fig. 4c). Likewise, truncation of the 5' end of the TLS to abrogate the linchpin interaction reduced binding approximately threefold ($K_d = 1.1 \pm 0.3$ nM) (Extended Data Fig. 5).

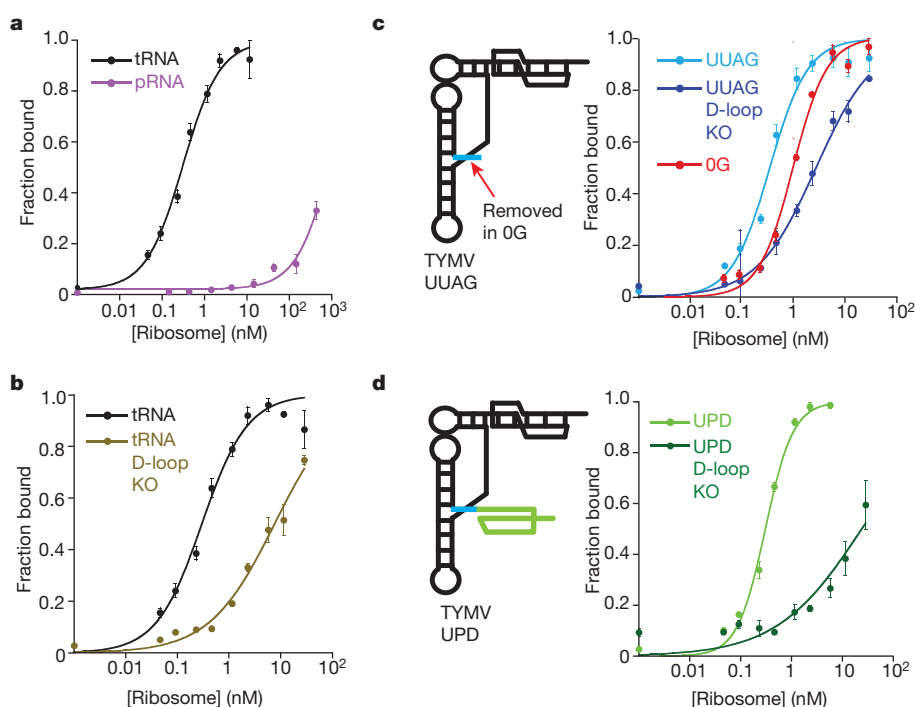


Figure 4 | Binding of tRNA and TLS to ribosomes. **a**, Binding curves of tRNA^{Val} (positive control) and pRNA (negative control) to 70S ribosomes, fit by a Langmuir isotherm (for RNA sequences see Extended Data Fig. 1). **b**, Binding of wild-type tRNA and tRNA with mutated D loop (D-loop knockout (KO)). **c**, Diagram of the UUAG TLS (UUAG sequence in cyan) and binding curves of this TLS and versions with the D loop mutated and with the UUAG removed (0G). **d**, Diagram of the UPD TLS (UPD shown in green) and binding curves of this UPD TLS and a D-loop mutant. Error bars are 1 standard deviation from mean of 3 replicates.

Remarkably, an RNA containing the TLS, the UUAG and the 23-nucleotide-long UPD also bound to ribosomes (TYMV UPD; $K_d = 0.24 \pm 0.11$ nM), and binding of this RNA was reduced 100-fold by D-loop mutation ($K_d > 24 \pm 8$ nM) (Fig. 4d). Thus, the folded TLS can bind the ribosome even in the context of the entire 3' UTR and binding depends on native structure. The affinity is consistent with binding to the P site, although binding to other sites is possible. The ability of the entire TYMV 3' UTR to dock within ribosomes may relate to its functions as a regulatory switch, a translation enhancer and a means to protect the 3' end of the genomic RNA^{17–19}.

METHODS SUMMARY

In vitro transcribed RNA was crystallized by vapour diffusion. Crystals grew to full size in 1–2 days, were derivatized with iridium (III) hexamine and cryo-protected. Diffraction data were collected at Advanced Light Source beamline 4.2.2 and used in single-wavelength anomalous dispersion (SAD) phasing. Crystal diffraction data, phasing and refinement statistics are contained in Extended Data Table 1. Ribosome binding was measured by filter binding.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 January; accepted 14 April 2014.

Published online 8 June 2014.

- Dethoff, E. A., Chugh, J., Mustoe, A. M. & Al-Hashimi, H. M. Functional complexity and regulation through RNA dynamics. *Nature* **482**, 322–330 (2012).
- Yot, P., Pinck, M., Haenni, A. L., Duranton, H. M. & Chapeville, F. Valine-specific tRNA-like structure in turnip yellow mosaic virus RNA. *Proc. Natl Acad. Sci. USA* **67**, 1345–1352 (1970).
- Pinck, M., Yot, P., Chapeville, F. & Duranton, H. M. Enzymatic binding of valine to the 3' end of TYMV-RNA. *Nature* **226**, 954–956 (1970).
- Litvak, S., Carr, D. S. & Chapeville, F. TYMV RNA as a substrate of the tRNA nucleotidyltransferase. *FEBS Lett.* **11**, 316–319 (1970).
- Hammond, J. A., Rambo, R. P. & Kieft, J. S. Multi-domain packing in the aminoacylatable 3' end of a plant viral RNA. *J. Mol. Biol.* **399**, 450–463 (2010).
- Dreher, T. W. Role of tRNA-like structures in controlling plant virus replication. *Virus Res.* **139**, 217–229 (2009).
- Giegé, R., Frugier, M. & Rudinger, J. tRNA mimics. *Curr. Opin. Struct. Biol.* **8**, 286–293 (1998).
- Cliffe, A. R., Nash, A. A. & Dutia, B. M. Selective uptake of small RNA molecules in the virion of murine gammaherpesvirus 68. *J. Virol.* **83**, 2321–2326 (2009).
- Dreher, T. W. Viral tRNAs and tRNA-like structures. *Wiley. Interdiscip. Rev. RNA* **1**, 402–414 (2010).
- Wilusz, J. E., Freier, S. M. & Spector, D. L. 3' End processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell* **135**, 919–932 (2008).
- Hammond, J. A., Rambo, R. P., Filbin, M. E. & Kieft, J. S. Comparison and functional implications of the 3D architectures of viral tRNA-like structures. *RNA* **15**, 294–307 (2009).
- Rietveld, K., Linschooten, K., Pleij, C. W. & Bosch, L. The three-dimensional folding of the tRNA-like structure of tobacco mosaic virus RNA. A new building principle applied twice. *EMBO J.* **3**, 2613–2619 (1984).
- Rietveld, K., Pleij, C. W. & Bosch, L. Three-dimensional models of the tRNA-like 3' termini of some plant viral RNAs. *EMBO J.* **2**, 1079–1085 (1983).
- Rietveld, K., Van Poelgeest, R., Pleij, C. W., Van Boom, J. H. & Bosch, L. The tRNA-like structure at the 3' terminus of turnip yellow mosaic virus RNA. Differences and similarities with canonical tRNA. *Nucleic Acids Res.* **10**, 1929–1946 (1982).
- Florentz, C. *et al.* The tRNA-like structure of turnip yellow mosaic virus RNA: structural organization of the last 159 nucleotides from the 3' OH terminus. *EMBO J.* **1**, 269–276 (1982).
- Giegé, R., Florentz, C. & Dreher, T. W. The TYMV tRNA-like structure. *Biochimie* **75**, 569–582 (1993).
- Matsuda, D. & Dreher, T. W. The tRNA-like structure of turnip yellow mosaic virus RNA is a 3'-translational enhancer. *Virology* **321**, 36–46 (2004).
- Singh, R. N. & Dreher, T. W. Turnip yellow mosaic virus RNA-dependent RNA polymerase: initiation of minus strand synthesis *in vitro*. *Virology* **233**, 430–439 (1997).
- Skuzeski, J. M., Bozarth, C. S. & Dreher, T. W. The turnip yellow mosaic virus tRNA-like structure cannot be replaced by generic tRNA-like elements or by heterologous 3' untranslated regions known to enhance mRNA expression and stability. *J. Virol.* **70**, 2107–2115 (1996).
- Matsuda, D., Yoshinari, S. & Dreher, T. W. eEF1A binding to aminoacylated viral RNA represses minus strand synthesis by TYMV RNA-dependent RNA polymerase. *Virology* **321**, 47–56 (2004).
- Shi, H. & Moore, P. B. The crystal structure of yeast phenylalanine tRNA at 1.93 Å resolution: a classic structure revisited. *RNA* **6**, 1091–1105 (2000).
- Dumas, P. *et al.* 3-D graphics modelling of the tRNA-like 3'-end of turnip yellow mosaic virus RNA: structural and functional implications. *J. Biomol. Struct. Dyn.* **4**, 707–728 (1987).
- Singh, R. N. & Dreher, T. W. Specific site selection in RNA resulting from a combination of nonspecific secondary structure and –CCR– boxes: initiation of minus strand synthesis by turnip yellow mosaic virus RNA-dependent RNA polymerase. *RNA* **4**, 1083–1095 (1998).
- Deiman, B. A., Koenen, A. K., Verlaan, P. W. & Pleij, C. W. Minimal template requirements for initiation of minus-strand synthesis *in vitro* by the RNA-dependent RNA polymerase of turnip yellow mosaic virus. *J. Virol.* **72**, 3965–3972 (1998).
- Fukai, S. *et al.* Structural basis for double-sieve discrimination of L-valine from L-isoleucine and L-threonine by the complex of tRNA^{Val} and valyl-tRNA synthetase. *Cell* **103**, 793–803 (2000).
- Dreher, T. W., Tsai, C. H., Florentz, C. & Giegé, R. Specific valylation of turnip yellow mosaic virus RNA by wheat germ valyl-tRNA synthetase determined by three anticodon loop nucleotides. *Biochemistry* **31**, 9183–9189 (1992).
- Dreher, T. W. & Goodwin, J. B. Transfer RNA mimicry among tymoviral genomic RNAs ranges from highly efficient to vestigial. *Nucleic Acids Res.* **26**, 4356–4364 (1998).
- Barends, S., Bink, H. H., van den Worm, S. H., Pleij, C. W. & Kraal, B. Entrapping ribosomes for viral translation: tRNA mimicry as a molecular Trojan horse. *Cell* **112**, 123–129 (2003).
- Berthelot, F., Bogdanovsky, D., Schapira, G. & Gros, F. Interchangeability of factors and tRNA's in bacterial and eukaryotic translation initiation systems. *Mol. Cell. Biochem.* **1**, 63–72 (1973).
- Leontis, N. B. & Westhof, E. Geometric nomenclature and classification of RNA base pairs. *RNA* **7**, 499–512 (2001).

Acknowledgements We thank H. Noller for the gift of 70S ribosomes. We thank I. Tinoco Jr, C. Musselman and T. Dreher for critical reading of this manuscript. The University of Colorado (UC) Denver X-ray Facility is supported by UC Cancer Center Support Grant P30CA046934. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences of the US Department of Energy under contract #DE-AC02-05CH11231. J.S.K. is supported by grants GM081346 and GM097333 from the National Institutes of Health and is an Early Career Scientist of the Howard Hughes Medical Institute.

Author Contributions J.A.H. and G.M.R. designed the crystallization RNAs and identified initial crystals. T.M.C. and D.A.C. improved and grew the crystals. Crystals were harvested by T.M.C., who also solved, built and refined the structure. J.C.N. collected and processed synchrotron diffraction data. G.M.R. conducted the ribosome binding experiments. J.S.K. provided overall supervision and guidance, and together with T.M.C. and D.A.C. wrote the manuscript.

Author Information Atomic coordinates and structure factor amplitudes have been deposited in the Protein Data Bank under accession number 4P5J. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.S.K. (jeffrey.kieft@ucdenver.edu).

METHODS

General procedures

Chemical reagents and synthetic DNA. General chemical reagents were all of molecular biology grade or higher. All aqueous solutions were made using diethylpyr-carbonate (DEPC)-treated milli-Q water and routinely filtered through 0.22 µm sterile filtration systems (Millipore). DNA primers were purchased from Integrated DNA Technologies and used without further purification. Nucleic acid concentrations were determined by monitoring a solution's absorbance at 260 nm using a Nanodrop UV-Vis spectrophotometer (Thermo). Iridium (III) hexammine was synthesized as previously described³¹.

RNA transcription. Double-stranded (ds)DNA templates for transcription were made by PCR using template plasmid DNA that contained the sequence of interest (plasmids made using standard mutagenesis methods). DNA from a 1 ml PCR reaction was used in a 5 ml *in vitro* transcription reaction with final concentrations of 30 mM Tris-HCl pH 8.0, 10 mM dithiothreitol (DTT), 0.1% Triton X-100, 0.1% Spermidine, 40 mM MgCl₂, 4 mM each NTP, and T7 RNA polymerase. The reaction was incubated at 37 °C for 6 h. Inorganic pyrophosphate was pelleted at 3,000g for 10 min, followed by EtOH precipitation of the supernatant. Precipitated RNA was pelleted by centrifugation, dried, then resuspended in 8 M urea. RNA was purified on a 10% denaturing PAGE slab gel at 40 W for 5 h, then excised and passively eluted in DEPC-treated water overnight at 4 °C. RNA was concentrated and exchanged into DEPC-treated water by ultrafiltration and stored at -20 °C.

RNA crystallization and diffraction data collection. The RNA sequence used in crystallization was based on a sequence identified by *in vitro* selection for TYMV TLS RNAs capable of efficient valylation and contained a point mutation in the AC loop³². This RNA was prepared for crystallography in a solution containing 5 mg ml⁻¹ RNA, 2.5 mM MgCl₂, and 10 mM HEPES-KOH pH 7.5. This mixture was heated to 65 °C for 3 min, then cooled at room temperature. After cooling, Spermidine was added to 0.5 mM. The reaction was centrifuged for 10 min at 13,000g and then used in sitting-drop vapour diffusion crystallization at 4 °C. One-microlitre of RNA solution was combined with 2 µl of 10% MPD, 40 mM Na-Cacodylate pH 6.0, 12 mM Spermine, 80 mM NaCl and 20 mM MgCl₂. The well solution was 20–35% MPD. Crystals appeared and grew to full size over the course of 1–2 days. To obtain derivatized crystals for phasing, a solution matching the well solutions with the addition of 8 mM iridium (III) hexammine was exchanged with the crystal growth solution. Crystals were harvested directly from the drops into nylon loops and flash-frozen by plunging into liquid nitrogen. Diffraction data were collected at Advanced Light Source beamline 4.2.2 using 'shutterless' collection at the iridium L-III edge (1.0972 Å) at 100 K. For each crystal, multiple 180° data sets were collected with 0.1° oscillation images. Data were indexed, integrated, and scaled using XDS^{33,34}.

Structure determination and refinement. Although data were collected and processed to 1.99 Å, only data to 2.5 Å were used for phasing. Fifteen iridium (III) hexammine sites were identified and used in single-wavelength anomalous dispersion (SAD) phasing within the AUTOSOL function of PHENIX (overall figure of merit (FOM) = 0.448; ref. 35). Scattering factors used were $f' = -11.92$, $f'' = 10.09$. Density modification using RESOLVE (solvent content set to ~50%) led to an interpretable electron density map (Extended Data Fig. 2). Iterative rounds of model building and refinement (simulated annealing, rigid-body, *B*-factor refinement, phase combination using COOT^{36,37} and PHENIX REFINE) led to the final model. The final model contains 84 of 86 nucleotides, 2 Mg²⁺ ions, 12 iridium (III) hexammine ions, one Spermine molecule and 126 water molecules. Crystal diffraction data, phasing, and refinement statistics are contained in Extended Data

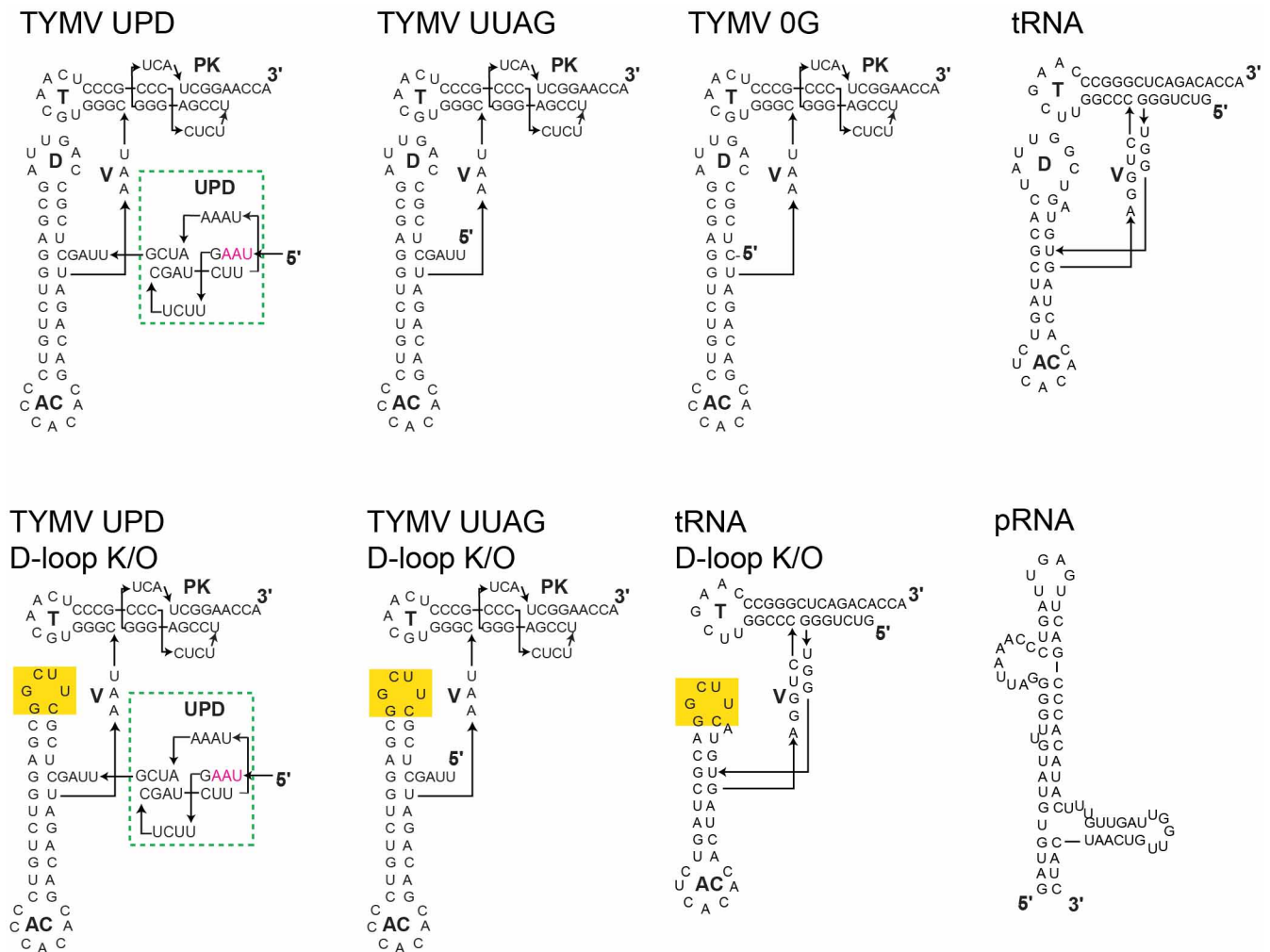
Table 1. Further analysis of the structure was completed using MolProbity^{38,39}. Summary of the output: clashscore = 14.19; probably wrong sugar puckers: 2; bad backbone conformations: 7; bad bonds: 0; bad angles: 3. Areas of concern were examined in the structure and generally fell within areas of the structure with unusual conformations, but the density and model agreed well in these regions.

Mutagenesis for ribosome binding. Mutations to the DNA templates were made using a PCR-based site-directed mutagenesis protocol (Agilent) with primers designed to modify the D-loop nucleotides. The nucleotides comprising the D loops of tRNA^{Val}, the TYMV UUAG TLS and the TYMV UPD TLS were replaced with stable UUCG tetraloop sequences. For tRNA, the primer sequence was 5'-GGG TGGTGTACTTCGGACGCTAGTCTC-3'. The UPD primer had the sequence 5'-CTTTAAATCGTTAGCTCGCTTCGGCGAGGTCTGTCCCC-3'. The UUAG primer sequence was 5'-CCGTCTTAGCTCGCTTCGGCGAGGTCTGTCCCC-3'.

70S ribosome purification. Preparation of 70S ribosomes was done by the Noller laboratory as previously described⁴⁰.

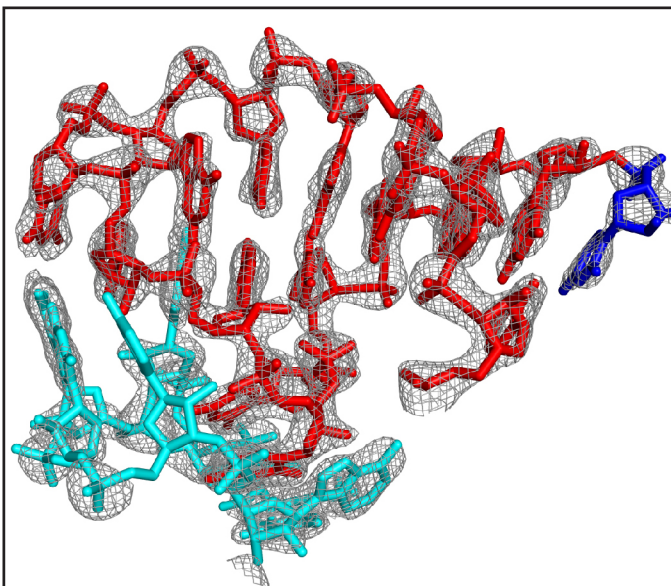
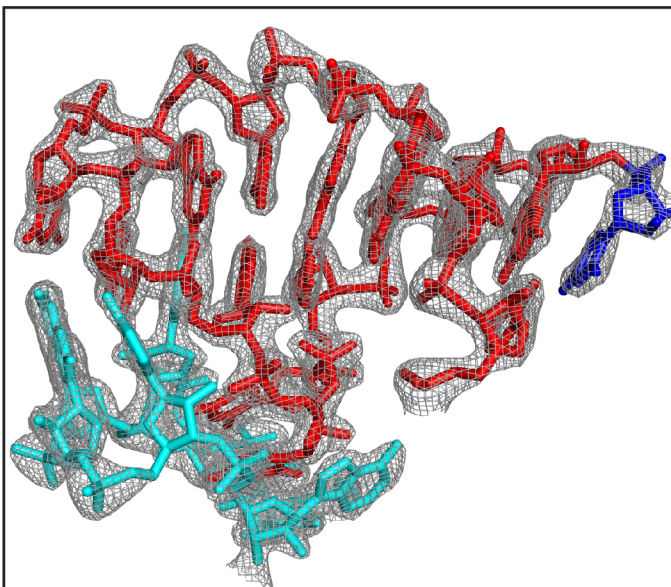
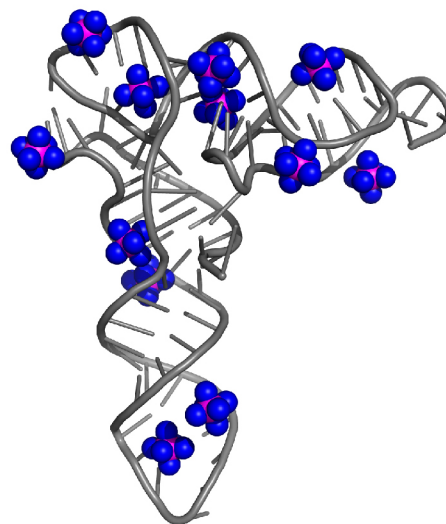
Filter binding. The filter binding protocol used was modified from previously published methods^{41,42}. Fifty-microlitre reactions contained 25 mM Tris-HCl, 50 mM KCl, 10 mM MgCl₂, 2 mM Spermine at pH 7.0, 100 counts per minute of ³²P-labelled RNA. The reactions were incubated at 37 °C for 30 min then passed through a sandwich of filters (pre-soaked in matching buffer) in a vacuum manifold. Filters: size exclusion (Tuffryn) filter (Pall), nitrocellulose filter (BioRad), Hybond-N+ charged nylon filter (GE BioSciences), and filter paper (Whatman). The filters were washed three times with wash buffer (25 mM Tris-HCl, 100 mM KCl, 25 mM MgCl₂, pH 7.5) and allowed to dry for 3 h. Reactions were quantified by phosphorimaging and data were fit using KaleidaGraph software.

- Keel, A. Y., Rambo, R. P., Batey, R. T. & Kieft, J. S. A general strategy to solve the phase problem in RNA crystallography. *Structure* **15**, 761–772 (2007).
- Wientges, J., Putz, J., Giege, R., Florentz, C. & Schwienhorst, A. Selection of viral RNA-derived tRNA-like structures with improved valylation activities. *Biochemistry* **39**, 6207–6218 (2000).
- Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).
- Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D* **66**, 133–144 (2010).
- Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
- Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
- Davis, I. W., Murray, L. W., Richardson, J. S. & Richardson, D. C. MolProbity: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res.* **32**, W615–W619 (2004).
- Zhu, J. et al. Crystal structures of complexes containing domains from two viral internal ribosome entry site (IRES) RNAs bound to the 70S ribosome. *Proc. Natl Acad. Sci. USA* **108**, 1839–1844 (2011).
- Costantino, D. & Kieft, J. S. A preformed compact ribosome-binding domain in the cricket paralysis-like virus IRES RNAs. *RNA* **11**, 332–343 (2005).
- Kieft, J. S., Zhou, K., Jubin, R. & Doudna, J. A. Mechanism of ribosome recruitment by hepatitis C IRES RNA. *RNA* **7**, 194–206 (2001).
- Kolk, M. H. et al. NMR structure of a classical pseudoknot: interplay of single- and double-stranded RNA. *Science* **280**, 434–438 (1998).
- Nissen, P. et al. Crystal structure of the ternary complex of Phe-tRNAPhe, EF-Tu, and a GTP analog. *Science* **270**, 1464–1472 (1995).
- Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030–1033 (2012).



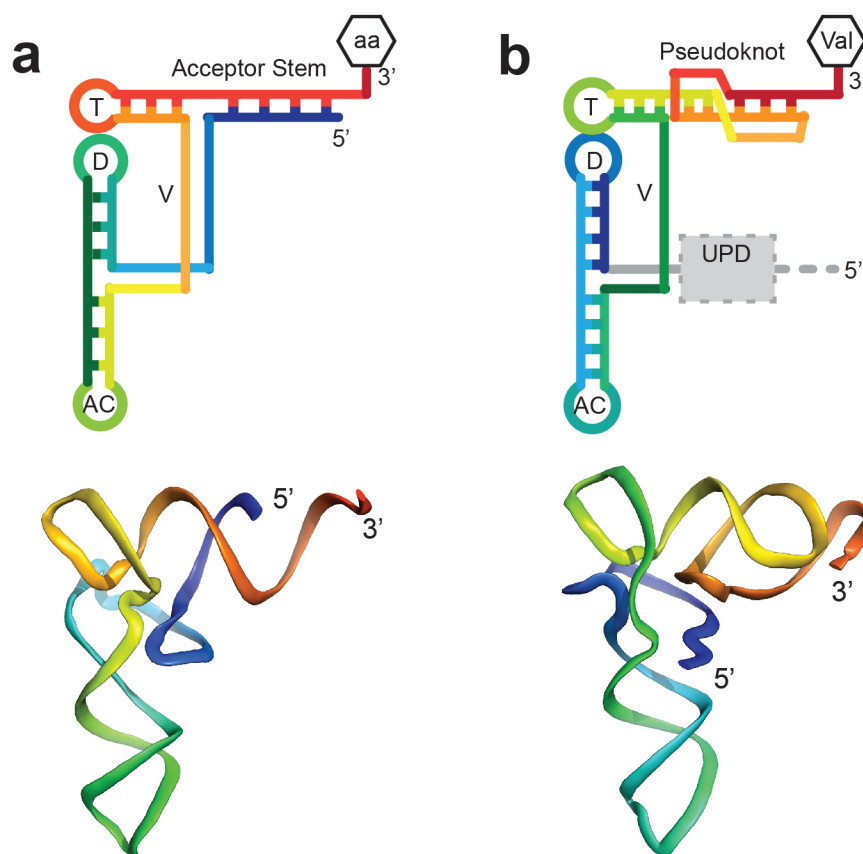
Extended Data Figure 1 | Sequences and structures of RNAs. Top left, sequence and secondary structure of the complete TYMV TLS and the UPD (green dashed box). The UPD is just upstream of the UUAG sequence that is important for stabilizing the L-shaped structure and the UPD is known to be

able to pack against the TLS^{5,11}. Interestingly, the stop codon for the Coat protein is within the UPD (magenta). Right and bottom, sequences and secondary structures of all additional RNAs used in ribosome binding assays or discussed in the text. Yellow highlights indicate the location of mutation.

a**b****c**

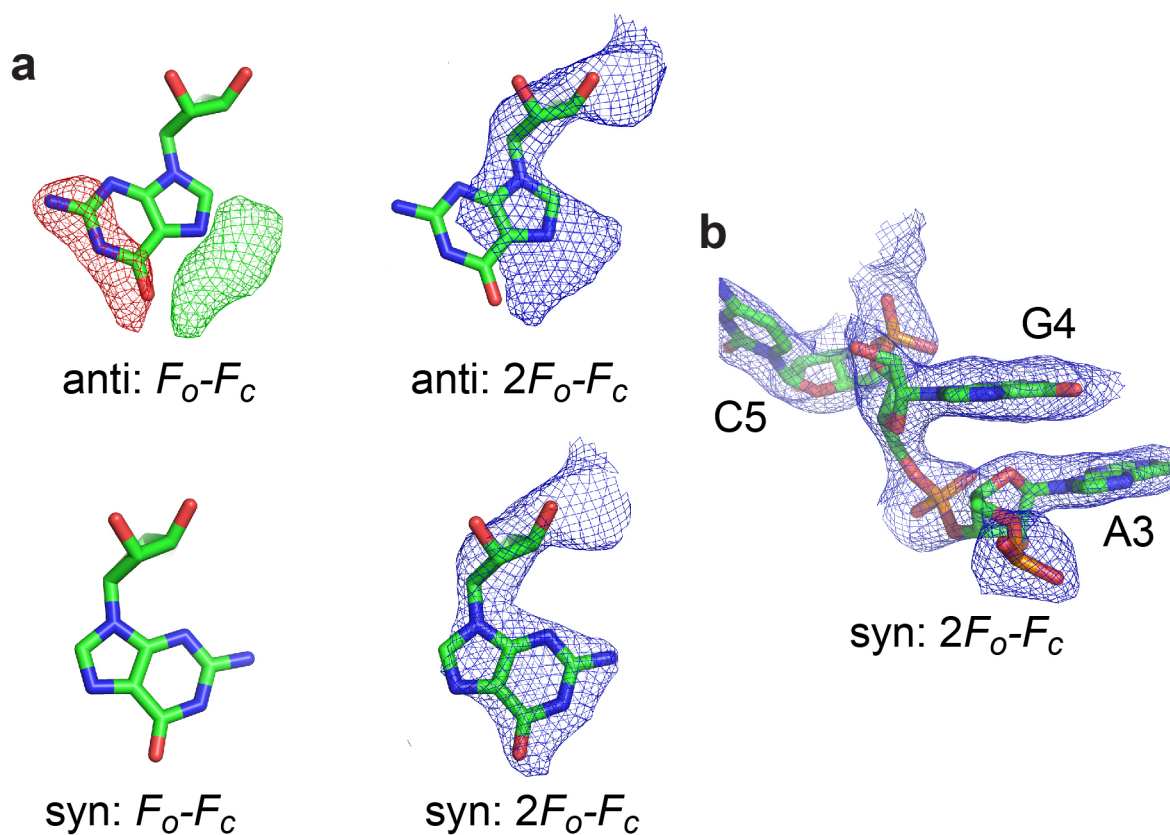
Extended Data Figure 2 | Representative electron density and bound trivalent ions. **a**, Unbiased, density-modified electron density from SAD phasing using data to 2.5 Å (grey mesh, 2σ), superimposed on the final model. The T loop and part of the D loop is shown. For simplicity, the density and structure of water and ions is not shown. **b**, Final $2F_o - F_c$ electron density map after model building and refinement to 1.99 Å (2σ). **c**, Structure with the location of 12 iridium (III) hexammine ions. Although many of these

hexammine binding sites may also be Mg^{2+} binding sites important for stabilizing the fold, the trivalent hexammine was present at 8 mM and thus many weaker Mg^{2+} binding sites could have been occupied. For this reason, and because there is not a one-to-one correlation of Mg^{2+} binding sites and trivalent hexammine sites, we do not make conclusions about Mg^{2+} binding on the basis of this structure.



Extended Data Figure 3 | Topologies and three-dimensional structures of tRNA and the TYMV TLS. **a**, Top, the topology of a canonical tRNA is shown in rainbow colours with the 5' end in blue and the 3' end in red. The attached amino acid is shown (labelled 'aa' or Val) and structural features are labelled: AC, anticodon loop; D, D loop; T, T loop; V, variable loop. The 5'

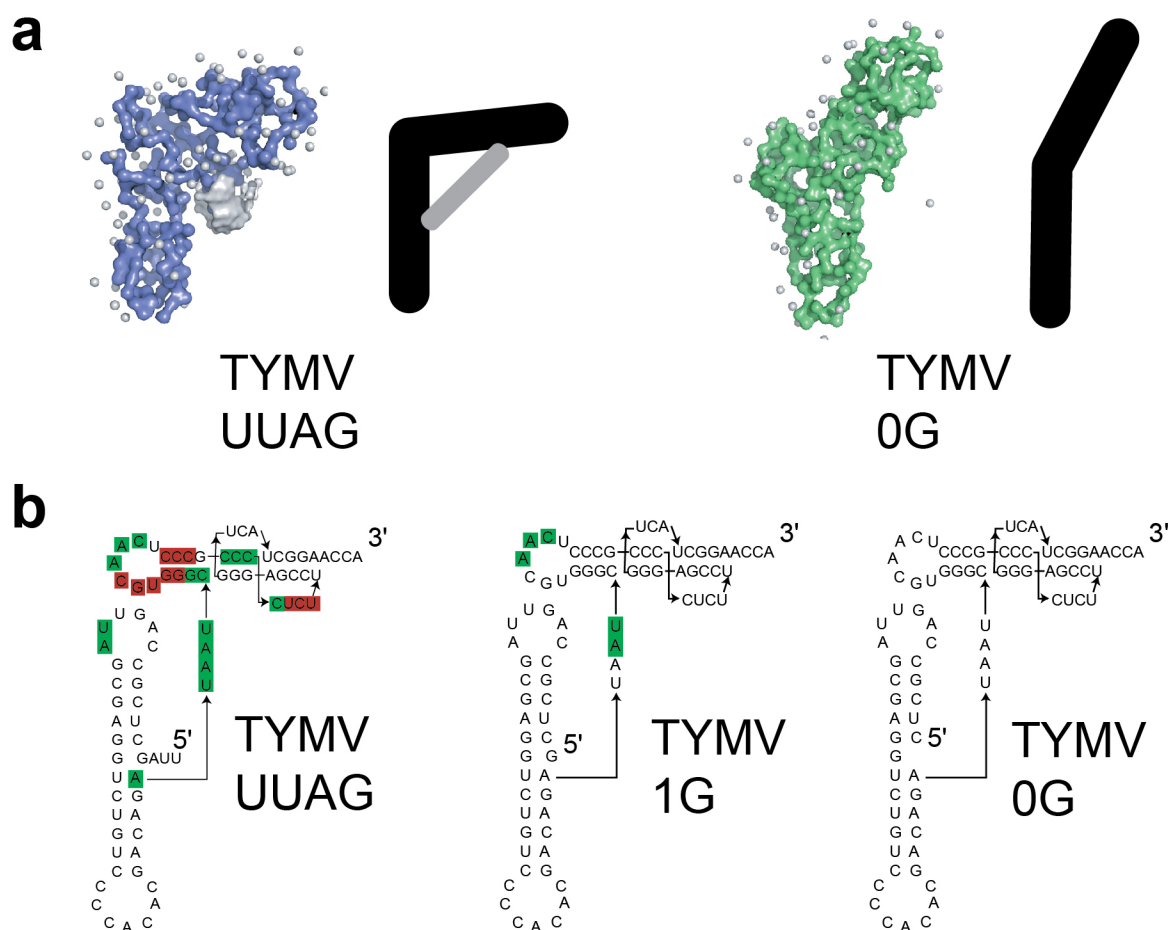
and 3' ends of the RNA are shown. Bottom, ribbon representation of the backbone of tRNA^{Phe} coloured roughly to match the cartoon diagram. **b**, Same as **a**, but for the TYMV TLS. The location of the UPD (grey dashed box) and gRNA (grey dashed line connected to the 5' end) are shown on the top diagram.



Extended Data Figure 4 | Assignment of bases to the *syn* conformation.

a, Nucleotide G4, which forms the long-range base pair with C76 in the pseudoknot, is in a *syn* conformation. Top, placement of the base into an *anti* conformation results in positive and negative density (green and red, respectively) in the $F_o - F_c$ map (left, contoured at 3σ), and the $2F_o - F_c$ map (right) shows the base is incorrectly placed (blue density, contoured at 1.5σ).

In contrast, placement of the base into the *syn* conformation (bottom) results in a flat $F_o - F_c$ map (left, contoured at 3σ) and a good fit to the $2F_o - F_c$ map (right, blue density contoured at 1.5σ). Base A11 is also in a *syn* conformation; the same analysis was performed to verify this (data not shown). **b**, $2F_o - F_c$ map surrounding bases A3–C5. The C4'–C5' bond of G4 is best modelled in the *trans* conformation.

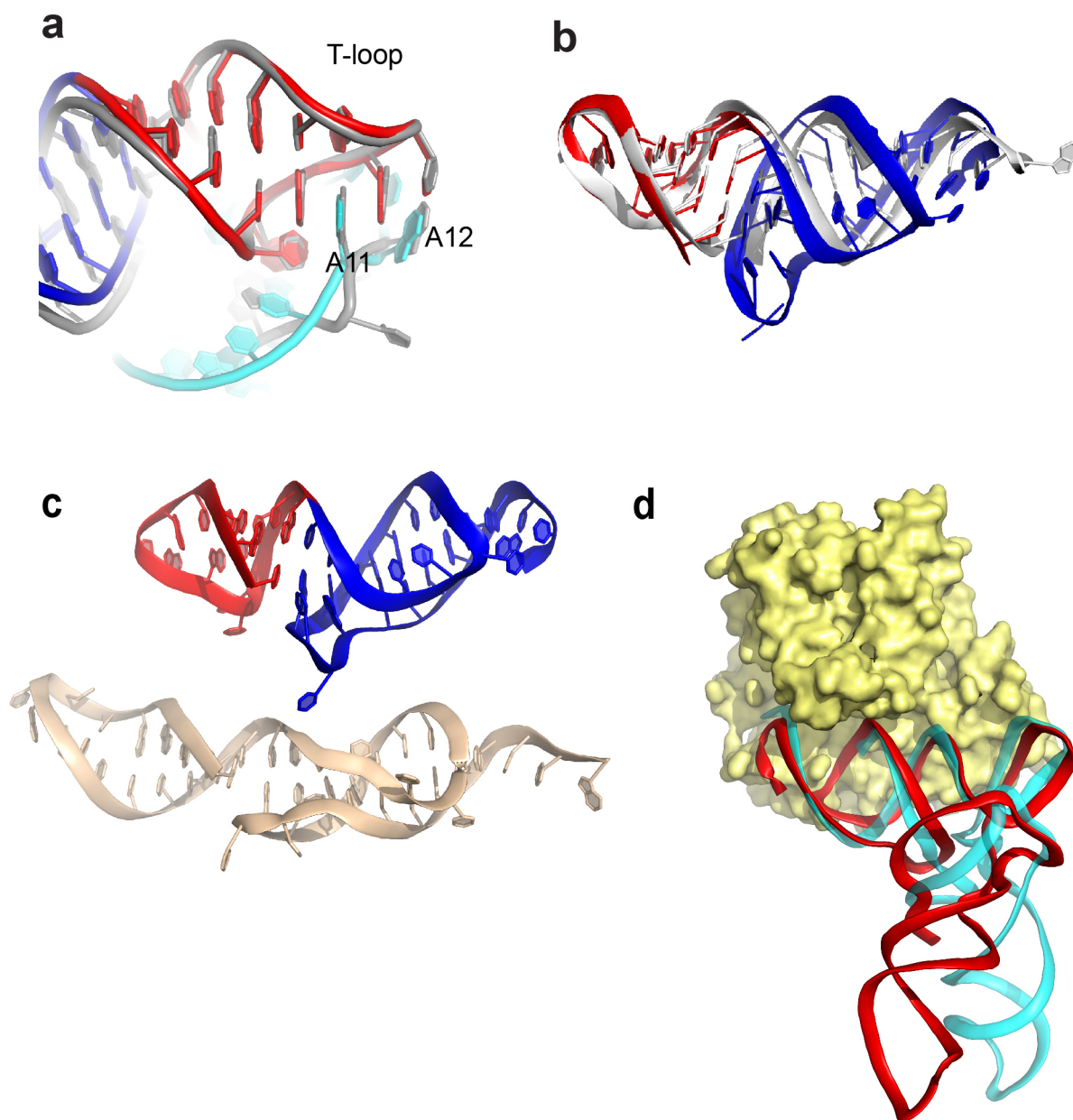


Extended Data Figure 5 | Effect of breaking the linchpin interaction.

a, Small-angle X-ray scattering (SAXS) analysis of TYMV TLS RNAs, adapted with permission from ref. 5. Left, *ab initio* SAXS reconstruction of the shape of the TLS when the 5' sequence that interacts with the pseudoknot (Fig. 2) is present. The RNA forms an L shape overall, illustrated by the black bars (stabilizing long-range interaction in grey). When these 5' nucleotides are removed (right), the L shape is lost and the RNA becomes more extended.

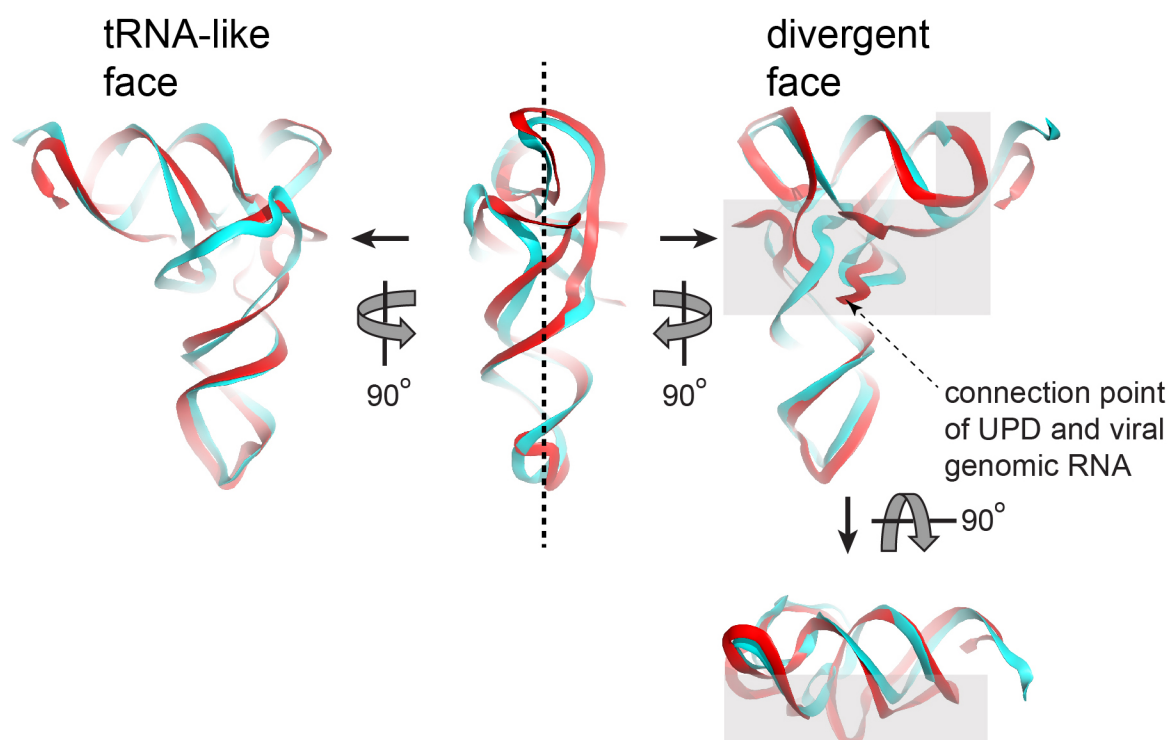
b, Hydroxyl radical probing of several TYMV TLS RNAs that indicate the effect of disrupting the long-range interaction, adapted with permission from refs 5,

11. Green and red indicate protection from cleavage by radicals and enhanced cleavage by radicals, respectively. Overall, the presence of green and red indicate tightly folded RNA. When the 5' nucleotides that form the long-range interaction are present, the RNA stably folds (TYMV UUAG, left). Removal of the 5' nucleotides destabilizes the fold (TYMV 0G, right). The presence of just G4 on the 5' end partially stabilizes the RNA fold (TYMV 1G, middle), confirming its importance in folding and also indicating that the nucleotides adjacent to G4 further stabilize the fold.



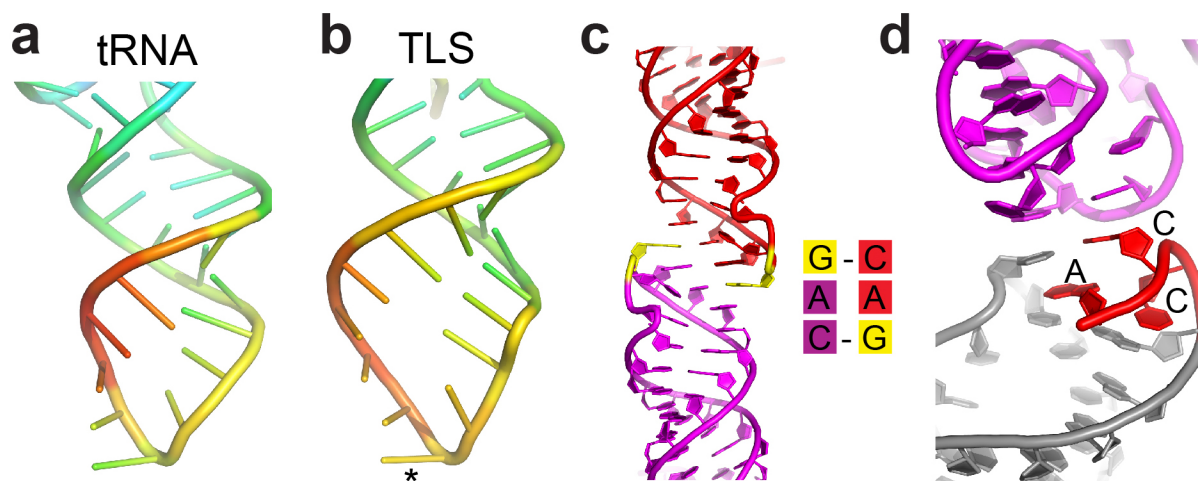
Extended Data Figure 6 | T loop and acceptor stems of the tRNA and TLS, and elongation factor binding. **a**, Superimposed structures of the TLS T loop (red) and part of the D loop (cyan) with the analogous structures in tRNA (grey). TLS bases A11 and A12 are shown; these bases match the interactions formed by analogous bases in tRNA. In the TLS, A11 is in a *syn* conformation, but the matching base in tRNA is not. This may be due to local differences in the backbone conformation. **b**, Superimposed structures of the TLS T loop (red) and pseudoknot (blue) with the T loop and acceptor stem elements in a tRNA (grey). View is from the 'top' of the molecule, down the axis of the D and AC

stems. **c**, Top, the structure of the T loop (red) and acceptor stem pseudoknot (blue) in the TLS crystal structure. Bottom, structure of these elements isolated from the rest of the TLS and solved by NMR (Protein Data Bank accession 1A60)⁴³. **d**, Superposition of the TLS structure (red) onto the tRNA (cyan) of a tRNA^{Phe} bound to EF-Tu (yellow), the bacterial homologue of eEF1A (Protein Data Bank accession 1TTT)⁴⁴. Binding is probably facilitated by the fact that the RNA backbone conformation of the TLS pseudoknot and T stem/loop matches that of a tRNA.



Extended Data Figure 7 | The 'two-faced' architecture of the TYMV TLS and connection with the UPD. Several views of the TLS (red) superimposed on tRNA^{Phe} (cyan)²¹ are shown, rotated 90° relative to each other. The dashed line bisects the structure into its two faces. The backbones are very similar

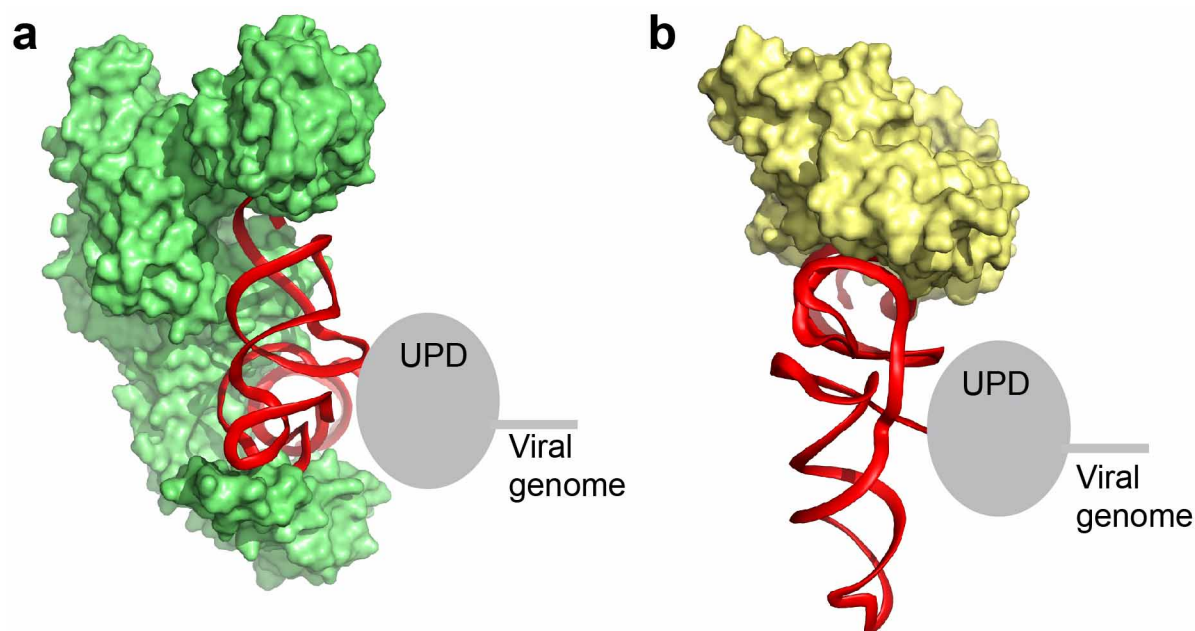
on the tRNA-like face, but differ on the divergent face. Locations where the two structures diverge most markedly are shaded grey. The 5' end of the TLS, where the UPD connects, is indicated.



Extended Data Figure 8 | The AC loop: structures and crystal packing.

a, Structure of the AC loop of tRNA^{Phe}, solved to 1.93 Å (ref. 21). The loop is coloured to reflect relative *B* factors, with red as the highest and blue as the lowest. **b**, Structure of the AC loop of the TYMV TLS, coloured identically to **a**. The asterisk marks the C30 base that was mutated to G to enhance crystallization. This was the only mutation made to the TLS for crystallization and does not inhibit aminoacylation³². Overall, the loop structures are similar and both have high crystallographic *B* factors compared with other parts of the structures, a common feature of tRNAs. There is no evidence that the TYMV TLS AC loop is post-transcriptionally modified, yet it has structural

features and conformational flexibility similar to the AC loop of a tRNA (which is often modified; Fig. 2a). **c**, Crystal packing involving the AC loop of the TYMV TLS. Two interacting copies of the RNA are shown in red and magenta, with the C30G mutation in yellow. This mutation, although not appearing to alter the overall AC-loop structure compared to a tRNA, induces intermolecular base pairing in the crystal (pattern shown to the right), suggesting why this mutation aided crystallization. **d**, Crystal packing of the 3' CCA of the TLS (red, labelled) against an adjacent molecule (magenta) probably causes the CCA to adopt a folded-back conformation.



Extended Data Figure 9 | Models of protein binding to the TLS and the location of the UPD. **a**, Model of the TLS (red, backbone ribbon shown) on the valine of AARS (green; Protein Data Bank accession 1GAX), similar to Fig. 3b, but viewed from the top and with the tRNA^{Val} not shown. The location of the UPD directly 5' of and against the TLS is shown as a grey oval. The viral genomic RNA is 5' of the UPD. Note that the strategy used by the TYMV TLS to interact with this protein is probably very different from that used by the

TLSs that are histidylated or tyrosylated, which are very different in terms of their secondary structure and fold^{6,9}. **b**, Same as **a**, but with the TLS modelled onto the bacterial homologue of eEF1A (EF-Tu) as in Extended Data Fig. 6. tRNA^{Phe} is not shown. In both complexes, the location of the 5' end, the UPD, and viral genome would not interfere with protein binding. This would not be true if the TLS had a tRNA-like topology with the 5' end paired to the 3' end.

Extended Data Table 1 | Crystallographic data collection, phasing and refinement statistics

Iridium (III) hexammine	
Data collection	
Space group	I 222
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	55.3, 101.6, 111.6
α , β , γ (°)	90, 90, 90
Resolution (Å)	28.87-1.99 (2.06-1.99)*
R_{sym} or R_{merge}	5.4 (82.3)
$R_{\text{meas}}^{\dagger}$	5.8 (89.5)
$I/\sigma I$	21.71 (2.19)
$CC(1/2)^{\ddagger}$	99.9 (83.3)
Completeness (%)	99.4 (94.7)
Redundancy	7.5 (6.5)
Refinement	
Resolution (Å)	28.9-1.99
No. reflections	308254 (18783)
$R_{\text{work}}/R_{\text{free}}$	20.6 (29.5) / 24.0 (33.3)
No. atoms	2011
RNA	1785
Ligand/ion	100
Water	126
B-factors	43.9
RNA	44.1
Ligand/ion	48.7
Water	38.5
R.m.s deviations	
Bond lengths (Å)	0.017
Bond angles (°)	1.83

One crystal was used.

*Highest-resolution shell is shown in parentheses.

$\dagger R_{\text{meas}}$ is R_{meas} as reported by XDS^{33,34}.

$\ddagger CC(1/2)$ is the percentage of correlation between intensities from random half-data sets as defined in ref. 45.

CORRIGENDUM

doi:10.1038/nature13569

Corrigendum: Sustained translational repression by eIF2 α -P mediates prion neurodegeneration

Julie A. Moreno, Helois Radford, Diego Peretti, Joern R. Steinert, Nicholas Verity, Maria Guerra Martin, Mark Halliday, Jason Morgan, David Dinsdale, Catherine A. Ortori, David A. Barrett, Pavel Tsytler, Anne Bertolotti, Anne E. Willis, Martin Bushell & Giovanna R. Mallucci

Nature **485**, 507–511 (2012); doi:10.1038/nature11058

It has been brought to our attention that there is an error in Supplementary Fig. 1b, owing to incorrect assembly of the image. The correct panel and figure legend (and the raw data used to generate Supplementary Fig. 1b) are shown in the Supplementary Information to this Corrigendum.

We would also like to clarify that the antibody used in Fig. 2e of the original Letter is Millipore mab1637. This antibody reacts with the C terminus of the neuron-specific β III isoform of β -tubulin. It does not identify β -tubulin in glial or non-neuronal cells, hence the absence of tubulin bands in control HeLa cells.

Supplementary Information is available in the online version of this Corrigendum.

ERRATUM

doi:10.1038/nature13571

Erratum: CTP synthase 1 deficiency in humans reveals its central role in lymphocyte proliferation

Emmanuel Martin, Noé Palmic, Sylvia Sanquer, Christelle Lenoir, Fabian Hauck, Cédric Mongellaz, Sylvie Fabrega, Patrick Nitschké, Mauro Degli Esposti, Jeremy Schwartzentruher, Naomi Taylor, Jacek Majewski, Nada Jabado, Robert F. Wynn, Capucine Picard, Alain Fischer, Peter D. Arkwright & Sylvain Latour

Nature **510**, 288–292 (2014); doi:10.1038/nature13386

Owing to a production error, the vertical axis of the right panel of Fig. 3g was misaligned. The correct panel is shown below as Fig. 1 of this Erratum. In addition, the legends for Fig. 2 and Extended Data Fig. 3 should read “Induction of CTPS1 expression during T- and B-cell activation and defective proliferation of activated CTPS1-deficient T and B cells” and “Induction of CTPS1 expression in activated B cells and inhibitors of CTPS1 expression in activated T cells”, respectively.

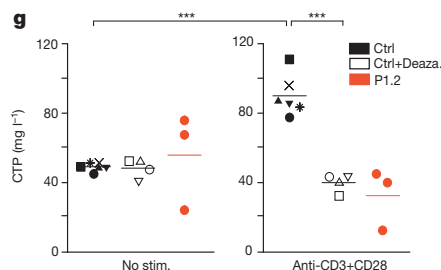


Figure 1 | This is the corrected Fig. 3g of the original Letter.

RETRACTION

doi:10.1038/nature13549

Retraction: Generation of cell polarity in plants links endocytosis, auxin distribution and cell fate decisions

Pankaj Dhonukshe, Hirokazu Tanaka, Tatsuaki Goh, Kazuo Ebine, Ari Pekka Mähönen, Kalika Prasad, Ikram Blilou, Niko Geldner, Jian Xu, Tomohiro Uemura, Joanne Chory, Takashi Ueda, Akihiko Nakano, Ben Scheres & Jiří Friml

Nature **456**, 962–966 (2008); doi:10.1038/nature07409

Our Letter reported that PIN transporters for the plant hormone auxin are initially delivered to the plasma membrane in a non-polar manner and that their polar distribution requires endocytosis. Abolishing PIN polarization, such as by inhibiting endocytosis, interferes with local auxin responses in the embryo, leading to transformation of embryonic leaves to the root-like structures.

The data regarding the essential role of endocytosis in the PIN polar localization and the connection between PIN polarity, auxin distribution and cell fate decisions remain reliable, but we have come to realize that the interpretation concerning the initial non-polar delivery of PIN proteins to the plasma membrane is not fully supported by experiments. It concerns the fluorescence recovery after photobleaching (FRAP) experiments presented in Fig. 1a and Supplementary Fig. 2a, which provided the key suggestion for the non-polar delivery model. On the basis of the original data, we confirm that the experiments were performed as published, but despite multiple attempts to reproduce them, the results remain inconclusive. Although in some experiments non-polar recovery can be detected as reported, others yield contrasting outputs that suggest polar recovery. Importantly, in many cases, the cells show signs of severe stress and stop growing following the photobleaching. In light of these findings, we feel that the reported results cannot be used for the conclusion on the initial non-polar PIN delivery and that this question remains open.

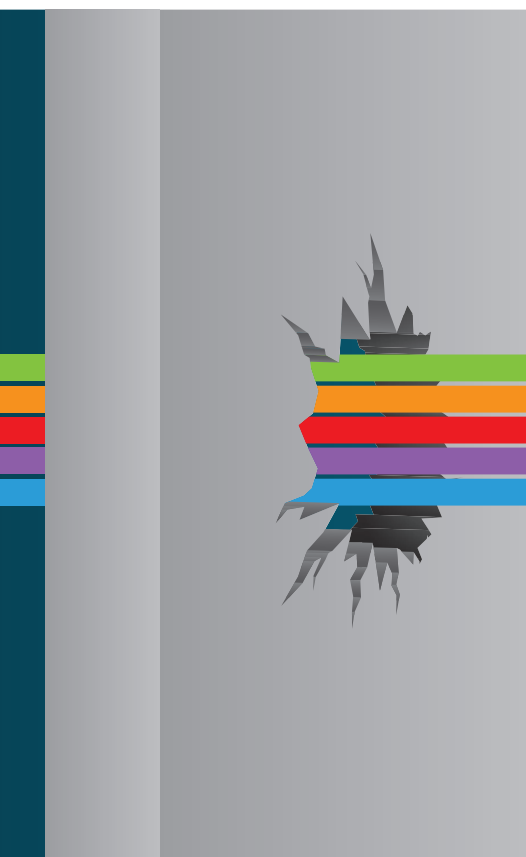
Therefore, we prefer to retract this Letter and republish the remaining confirmed findings elsewhere. Author P.D. continues to stand by all the conclusions of the paper, but all the other authors agree with the retraction. We apologise for any adverse consequences that may have resulted from this situation.

CAREERS

TURNING POINT Computer scientist models global ecology **p.373**

NATUREJOBS BLOG The latest on careers news and tips <http://blogs.nature.com/naturejobs>

NATUREJOBS For the latest career listings and advice www.naturejobs.com



INTERDISCIPLINARY RESEARCH

Break out

Researchers working at the interface of disciplines can pursue insights without sacrificing career progress.

BY VIRGINIA GEWIN

Compostable electronics, bacterial communication and forced-migration prediction. These seemingly unrelated research topics have underlying similarities: they are all examples of solution-oriented projects that require a broad cross-section of expertise.

Interdisciplinary research is starting to attract more and more attention — and funding. This year, for example, the US National Science Foundation (NSF) has requested US\$63 million (210% more than in 2012)

for its INSPIRE (Integrated NSF Support Promoting Interdisciplinary Research and Education) awards programme, which supports research into complex scientific problems such as space-weather monitoring, groundwater restoration and epigenomic analysis of single cells. In an era of stagnant, even shrinking, research funds, such budding fields can be a shrewd choice, especially for early-career researchers.

Interdisciplinary research pulls together disparate expertise to advance an emerging field or solve a multifaceted problem. Nanotechnology, for example, requires knowledge

of chemistry, biology and physics, and disease control can involve molecular biologists, biostatisticians, public-health officials and sociologists. Environmental science, with its study of entangled ecosystems and policy impacts, is the quintessential interdisciplinary field.

And, like the United States, the UK government has now dedicated funds to interdisciplinary research (see 'Interdisciplinary aid'). Research Councils UK (RCUK), the partnership of all seven publicly funded UK research councils, has identified six priority research areas, including energy and global food security, and in 2012 it joined with funding agencies from other nations in a €20-million (US\$27-million) initiative to support research on multinational, multidisciplinary problems, such as coastal vulnerability and freshwater security.

Funding agencies are not the only organizations to encourage young scientists into discipline-spanning research. Universities are making structural changes to promote and accommodate interdisciplinary research, most notably by creating interdisciplinary centres or institutes.

Last year, Stanford University in California launched a neuroscience institute and one for chemical biology, bringing the number of interdisciplinary laboratories, centres and institutes to 18. Heriot-Watt University in Edinburgh, UK, restructured in 2012 into 9 engineering and science institutes and 20 multidisciplinary centres, for areas such as sustainable-building design; sensors, signals and systems; and ocean systems.

MEASURE OF METRICS

But interdisciplinary research can have downsides. Perhaps counter-intuitively, interdisciplinary researchers must carve out a speciality, to form a coherent body of work from disparate strands. This can be difficult if the goal is innovation rather than getting work published, and evaluation metrics can be a major pitfall. Publications in high-profile journals are still the main scorecards for tenure and promotion decisions in many countries. In the United Kingdom, high-profile publications submitted from each researcher are the basis of comparison for the Research Excellence Framework, a government assessment that takes place every six years and establishes university funding levels. Because the evaluation weights ►

IVCANDY/ISTOCK VECTORS/GETTY

► research outputs (in the form of high-profile publications) at 65% of the total score, interdisciplinary research is liable to garner fewer funds under these discipline-focused standards.

The result is a large gap between the growing number of incentives to conduct interdisciplinary research and the level of career advancement it can offer. Even securing a junior interdisciplinary post is fraught with difficulty (see *Nature* **476**, 115–117; 2011), and career advancement for non-traditional research output poses even more challenges.

Stephanie Pfirman, a polar researcher with a joint appointment at Barnard College and Columbia University in New York city, can attest to the difficulties. Her interests in policy issues were stymied early in her career when her mentors and others consistently advised her to defer her ideas until she was more established. Given the challenges of becoming too broad too soon, she thinks it was good advice, but her experience led her to craft a report entitled *Interdisciplinary Hiring and Career Development: Guidance for Individuals and Institutions*, which was published in 2011 through the National Council for Science and the Environment (see ‘Quick tips’).

GROW STRONG AND BROAD

Interdisciplinary researchers have varied interests, so their work is published in journals that span different and sometimes unrelated disciplines. And yet academics are most likely to advance when their expertise is easily identifiable. What to do? Pfirman says that interdisciplinary academics should think of themselves as a tree: a researcher needs to have a main trunk of ideas, but also put out roots and branches that can connect to others. She points to Solomon Hsiang, an environmental scientist at the University of California, Berkeley, who was a lead author of *Risky Business*:

The Economic Risks of Climate Change in the United States, a high-profile report released in June. Hsiang combines large, independent sets of social-science, meteorological and climatological data with statistical methods that are more commonly used in microeconomics than in natural science. “I can give the same talk using the same data to climate scientists and to microeconomists, but they will look very different,” he says. He reckons that he almost did the work of two PhDs to get to the point of being a go-between in these fields. But it paid off in that it enabled him to do innovative research.

“One of the challenges of conducting interdisciplinary work is that it’s not always obvious which academic department is the best fit,” says Simon Goring, a postdoctoral researcher studying palaeoecology at the University of Wisconsin–Madison. He, for example, is based in the geography department, but says that his research on continental-scale ecological patterns could just as easily be classed as biology.

A home department will want to evaluate someone on the basis of their contributions to the core discipline of that department, says Laura Meagher, senior partner at Technology Development Group, a company based in Fife, UK, that advises higher-education institutions and research agencies on how to make strategic changes. Meagher once interviewed a department chair about the interdisciplinary postdoctoral fellows coming

through his department. The chair raved about the quality of their research and ideas, but admitted that he probably wouldn’t hire one because he needs faculty members who can teach the department’s introductory courses.

Perhaps the biggest professional concern for an interdisciplinary researcher is whether discipline-based tenure committees can adequately evaluate the impact of their work. Such committees often place substantial weight on external letters of support from knowledgeable faculty members at other institutions. Priyamvada Natarajan, an astronomer at Yale University in New Haven, Connecticut, describes how the dilemma could play out for a biophysicist who works on fluid flow, for example. It is possible that letter writers end up being biologists who may not fully appreciate the research contributions, rather than physicists who work on fluid flows, she says. “Those types of situations can result in tenure surprises.”

SET PRIORITIES

One way to ensure career advancement is to understand how university administrators measure success. Michael Binford, who studies land-use change at the University of Florida in Gainesville, notes that interdisciplinary researchers in productive groups should end up with more papers. But to ensure enough lead authorships, he says, they should carve out their own contribution and turn that into a paper.

Tenure committees may overlook some outputs of an interdisciplinary researcher, especially if it is a computer program or statistical tool. In a paper published in February (S. J. Goring *et al.* *Front. Ecol. Environ.* **12**, 39–47; 2014), Goring and his colleagues raised their concerns about being evaluated according to traditional success metrics, and said that tenure committees should be encouraged to value data-set creation, blogs, social media and policy-relevant activities.

Meagher emphasizes that it is essential for interdisciplinary scientists to highlight their unique abilities. “Don’t be afraid to say that you enable people to work together across teams and perspectives,” she says, adding that the teams that win interdisciplinary grants are the ones whose members make it clear in the proposal that they will spend time building trust and learning each other’s languages.

Researchers who seek academic employment should evaluate an institution’s track record of supporting and valuing interdisciplinary research before they accept a position there, says David Hassenzahl, newly appointed dean of California State University’s college of natural sciences in Chico. He points, for example, to a tool called STARS (Sustainability, Tracking, Assessment and Rating System; <https://stars.aashe.org>), which was put together by the Association



“If you try to be everywhere, you may not get traction anywhere.”

Stephanie Pfirman

INTERDISCIPLINARY AID

Funders and institutions show their support

The University of Southern California in Los Angeles is one of the few US institutions that has amended its promotion and tenure guidelines for interdisciplinary faculty members. In 2011, it allowed evaluation committees to consider letters of support from a mix of departments, and last year, it provided guidelines for assessing academic output beyond journal articles — including enhanced data sets, software and collaborative tools.

Funders are also encouraging interdisciplinary collaboration. Scottish Crucible, a scheme launched in 2009 and financed in part by the Scottish Funding

Council, provides a three-month leadership and communication training programme for 30 participants selected from wide-ranging disciplines.

The scheme offers a sort of ‘speed dating’ venue for participants, encouraging them to share their work with one another and to pursue potentially interesting partnerships.

At the end, participants can submit proposals for collaborative projects.

Versions of Scottish Crucible are popping up elsewhere in the United Kingdom: there is now a Welsh Crucible and a South West Crucible. **V.G.**

QUICK TIPS

Navigating disciplines

Stephanie Pfirman, a polar researcher at Columbia University in New York city who has published a report on how to handle interdisciplinary-research issues, suggests four ways that can help researchers to carve out their niche.

- Make sure that your CV spells out your contribution and how it was integral to the overall project.
- Attend the most relevant meeting in your core discipline and run a session on your topic to highlight its importance and to help spur connections.
- Indicate your academic reach. Include a link to your Google Analytics web page, for example, and use it as a citation index because it illustrates impacts more broadly.
- Expand your network by making contact with authors of papers that cite yours. **V.G.**

for the Advancement of Sustainability in Higher Education in Denver, Colorado, to identify universities and colleges that give interdisciplinary research the same weight as traditional disciplinary research.

There is also no escaping the fact that interdisciplinary research spans not just different disciplines but different academic cultures. Researchers often end up in joint appointments — a faculty position that reports to two departments. Such positions can be risky because the person is effectively serving two masters, who may have differing views on the achievements needed for tenure.

One strategy is to cultivate a large network in both fields, but some researchers caution against spreading their efforts too thinly. “If you try to be everywhere, you may not get traction anywhere,” says Pfirman. That said, she adds, visibility and recognition are crucial for early-career scholars. Young scientists should focus their efforts on a big disciplinary meeting that is close to their interdisciplinary speciality, she says, and make a name for themselves there.

Ultimately, any career arc needs to tell a coherent story. Early-career researchers need to make clear how their work ties together into a “meaningful, original research agenda,” says Hassenzhl. “That’s what academia is all about.” ■

Virginia Gewin is a freelance writer in Portland, Oregon.

TURNING POINT

Drew Purves

Drew Purves, who heads the Computational Ecology and Environmental Science group at Microsoft Research Cambridge, UK, published the first-ever mechanistic general ecosystem model in April (M. B. J. Harfoot et al. PLoS Biol. 12, e1001841; 2014). This tool simulates the interactions of all organisms on Earth and the underlying ecological mechanisms that govern biodiversity patterns, which may help to predict how invasive species or pollution shape the world.

How did you tackle the ecosystem model?

Stephen Emmott, our head of computational science, likes to take a broad-sweep approach to science. One day he asked: “Why don’t we model all life on Earth?” I was sceptical, but I like a challenge, and we wanted to do something that would be useful for the conservation community. It took four years, and the model turned out to be unusual — we couldn’t model every individual species, so the key development was figuring out how to properly simulate nature using realistic and rigorous approximations.

Are you a geek at heart?

Yes. I got into computer programming when I was 7 — I got a Commodore 64, one of the first home computers. Around age 14, I watched a documentary about artificial life, and started reading about how to simulate life through a computer. Looking back, I realize now that my interest as a student in examining real-life processes of ecology and evolution at the University of Cambridge evolved from my interest in studying artificial life.

How did your postdoc shape your career?

In 2001, I was lucky enough to get a postdoc with ecologist and evolutionary biologist Stephen Pacala at Princeton University in New Jersey. He had just become director of the Princeton Environmental Institute, which had support from the global oil company BP and the US car manufacturer Ford. As a result of working with him, I met several senior executives from these companies who wanted to do cool and risky stuff, such as carbon capture and storage. After meeting them, I was more open to considering the Microsoft research job when it came up. Before I had those experiences, I had presumed that big corporations were evil and only out for profit.

What convinced you to go for the Microsoft job?

The job advertisement sounded ambitious. They said they wanted a computational ecologist, a phrase I had never heard before, but it



MICROSOFT RESEARCH LTD

sounded like what I wanted to do. In many ways it didn’t make sense to take this position when I was starting to get offers to do the academic jobs I had trained for. But I rationalized that there would always be university jobs available.

How does this position differ from an academic research job?

There is no predefined idea of success here. There is an expectation that what we are doing will have an impact on society, but that impact could take the shape of high-profile publications or software development that could enhance the field of computational ecology. It makes me weep that in academia we take the cleverest people in society and rank them on a single dimension — their publication record. At Microsoft, we do not have to pursue predefined ideas. I can follow my interests — helping humanity to achieve a better understanding of nature and the biosphere we all depend on. My group is a small team with limited resources, but we take on big projects, such as predictive modelling of global agriculture.

Where do you go from here?

I want to run more scenarios to see how well real-world data fit our model and to try, for example, to predict outcomes under different climate conditions. I’ll use the results to explore interesting applied questions as well; for example, I would like to simulate Australia’s cane-toad invasion. We need to find ways to sufficiently connect our models to existing data; with enough of those links, we can put realistic limits on the model to learn how and where it works best. In my darkest moments, I wonder whether this is still science. But surely it’s science in the same way that we model how galaxies formed? ■

INTERVIEW BY VIRGINIA GEWIN

ARE YOU RECEIVING?

Out in the cold.

BY REBECCA BIRCH

Galactic Standard Date: 11657.3

Planetfall successful. Atmosphere breathable, as anticipated from earlier analysis. Base establishment under way, following standard protocol. Work is slow, given we're a five-person crew, but no unanticipated challenges yet reported.

Landscape is surreal. Frozen drifts and billows, like snow back home, but when you look just off of straight there are rainbow spectra dancing in the crystals. Winds are constant. Science Tech O'Malley reported hearing voices when she went outside to set up the solar panels, but the doctor assures me it's just the change in aural input after so long aboard ship. I'm confident initial planetary analysis showing no sign of intelligent life was accurate.

Captain Marjorie Halstone, awaiting confirmation of transmission.

Galactic Standard Date: 11663.8

Base operational, but not optimal. Solar-energy collectors hampered by constant snow accumulation. Panels have been reoriented to discourage build-up, and shifts have been instituted to clear off what does pile up. We've begun local reconnaissance on foot. Until proper energy levels are established, use of mechanized transport is unfeasible. The snow's spectral light phenomenon seems to intensify during night-time hours. Still awaiting confirmation of original transmission. Are you receiving?

Galactic Standard Date: 11672.5

Despite reorientation of panels, snow accumulation has not decreased, and panel surfaces are sustaining damage. This snow has abrasive properties not previously anticipated. Energy reserves are now below 60% of recommended. O'Malley continues to report hearing things and is no longer permitted alone surface-side, after attempting to follow the sounds out of range of communications. The doctor has prescribed sensitivity dampeners.

I have not told anyone about the sounds I hear on my own panel-clearing shifts. I prefer to remain unmedicated.

Reconnaissance has been curtailed for the moment to focus on snow abatement. Techs Akira and Butler are working to find a reliable

countermeasure, but as yet have had no success. Study of atmospheric data shows no sign of any foreseeable change in weather patterns. If no solution is found, I'm afraid I'll be forced to order the termination of this mission.

Stand ready to initiate evacuation procedures and please send immediate confirmation of all transmissions.

Galactic Standard Date: 11677.2

Butler is gone.

We didn't know he was missing until he failed to return from his night-time clearing shift. I attempted to track him, but the colours in the snow hid any footprints, and the farther I got from base ... Well, up to this point, I believed the sounds I was hearing were environmental, but now I swear there are words ...

Belay that last bit. No, Doctor, I don't require any dampeners. See to O'Malley. She and Butler were close. Please shut the door behind you.

Energy reserves have dipped below 40%. O'Malley is begging to go after Butler, even with an increased dosage of dampeners. The doctor has been drafted into panel maintenance, over his objections. We can't risk letting O'Malley outside again.

Captain Halstone requesting immediate evacuation. Before we lose another.

Galactic Standard Date: 11680.2

Dampeners weren't enough. This morning, O'Malley vanished. Left during my shift and I never saw her. Never heard her. Just those damn lights. I see them on the backs of my eyelids whenever I close them. Akira says he hasn't had more than three hours of sleep in the past two days. I'm not much better off. As for the doctor, he won't talk about the lights. Won't talk about anything. I saw him dosing himself with dampeners, though he claims he doesn't hear the voices.

Power reserves at 15%, well below emergency levels. Both Akira and I have triple-checked communication mechanics. Everything is in working order. Why aren't you responding? Send help now. Please.

Galactic Standard Date: 11682.2

Found the doctor dead in his bunk this morning of apparent dampener overdose. Energy reserves at 3% and falling. The cryochamber won't last once they power's gone,

so we've buried him in the snow just outside the exterior hatch. His family would wish to have his remains, if anyone should hear this message.

Akira thinks the voices may be originating from a point southwest of base. Remaining here is no longer an option. If there's something else alive out there and we can find it, then maybe we have a chance.

This will be the last communication.



We stagger together through a changed landscape. The snow-light is no longer a mosaic of scattered crystal prisms. Instead, a bright rainbow band spreads across the drifts, leading us southwest. I wouldn't believe it if Akira didn't see it, too. Our feet sink in with each step down the golden path in the centre, and we cling to each other for support.

The voices are clear now, rising up out of the snow. *Captain Halstone, abort landing. Unexplained phenomenon detected planet-side. Repeat, abort landing. Please confirm.*

My own voice, like a dream. *Awaiting confirmation of transmission ... Are you receiving?*

I hear O'Malley, too, and Butler. *You're almost here, Captain. Just a little farther. Akira, we're so glad you're coming.*

Just ahead, the rainbow narrows until it vanishes in a pool of silver light. Two familiar forms stand with arms outstretched, their bodies rimmed with kaleidoscopic auras.

Akira squeezes my arm. We head for the light. I don't know what's on the other side, and I don't know if we'll ever return, but I'm telling the wind our tale, hoping it will sing until someone comes after us. Someone who can bring the story home.

Ready, Akira? Let's go. ■

Rebecca Birch lives in Seattle and has been published in markets including the Grantville Gazette, Abyss & Apex and Penumbra. Find her online at www.wordsofbirch.com.

JACEY

➤ NATURE.COM

Follow Futures:

@NatureFutures

go.nature.com/mtoodm